

## Tesis de Maestría

# Caracterización de la confianza en las instituciones del gobierno del Perú

Arcaya Arhuata, Ludmer Edward

2016-07-12

Este documento forma parte de la colección de tesis doctorales y de maestría de la Biblioteca Central Dr. Luis Federico Leloir, disponible en [digital.bl.fcen.uba.ar](http://digital.bl.fcen.uba.ar). Su utilización debe ser acompañada por la cita bibliográfica con reconocimiento de la fuente.

This document is part of the doctoral theses collection of the Central Library Dr. Luis Federico Leloir, available in [digital.bl.fcen.uba.ar](http://digital.bl.fcen.uba.ar). It should be used accompanied by the corresponding citation acknowledging the source.

#### Cita tipo APA:

Arcaya Arhuata, Ludmer Edward. (2016-07-12). Caracterización de la confianza en las instituciones del gobierno del Perú. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.

#### Cita tipo Chicago:

Arcaya Arhuata, Ludmer Edward. "Caracterización de la confianza en las instituciones del gobierno del Perú". Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. 2016-07-12.

**EXACTAS** UBA

Facultad de Ciencias Exactas y Naturales



**UBA**

Universidad de Buenos Aires



## **Maestría en Explotación de Datos y Descubrimiento del Conocimiento**

Facultad de Ciencias Exactas y Naturales  
Facultad de Ingeniería  
Universidad de Buenos Aires

# **CARACTERIZACIÓN DE LA CONFIANZA EN LAS INSTITUCIONES DEL GOBIERNO DEL PERÚ**

Tesis presentada para obtener el título de  
Magister en Explotación de Datos y descubrimiento del Conocimiento

Ing. Ludmer E. Arcaya Arhuata

Directora: **Mg. Cecilia Ruz**

Buenos Aires 2016



## **Dedicatoria.**

A Dios todo poderoso, por la salud, día a día me permite llegar a cumplir esta meta tan importante.

A mis padres que siempre son mi motivación, mis hermanas que sin querer ellas llenan mi vida de alegría, mis profesores que han dado de sí para transmitir sus grandes conocimientos.

No tengo palabras para agradecer al Gobierno Peruano, por su apoyo en apostar por mí, por la beca integral sobre esta maestría.

A Cecilia por la dirección de esta tesis, que sin ella no habría sido posible ordenar mis ideas y concretar este trabajo.

A Todos los profesores de la Maestría que con su experiencia y conocimiento nos enseñaron y ampliaron un nuevo horizonte en mi vida profesional.

# Índice

<b>1. Introducción</b>	<b>7</b>
1.1. Planteamiento del Problema	8
1.2. Justificación del Problema	9
1.3. Antecedentes del Problema	9
1.4. Alcances y Limitaciones	10
1.5. Objetivo de Investigación	11
<b>2. Marco Teórico</b>	<b>12</b>
2.1. Estado del arte	12
2.2. Confianza Ciudadana	15
2.3. Las Instituciones Públicas del Perú	16
2.4. Encuesta de Hogares	20
2.5. Minería de Datos	21
2.6. Chi-Cuadrado	23
2.7. Análisis Factorial de Correspondencias	24
2.8. Arboles de Decisión	25
<b>3. Desarrollo</b>	<b>29</b>
3.1. Herramientas de Análisis	29
3.2. Análisis de la estructura de los datos	30
3.3. Etapa de Extracción Transformación y Carga(ETL)	33
3.4. Proceso de Exploración de datos y selección de variables	34
3.5. Selección de las variables más influyentes	43
3.6. Comparativa de comportamiento de variables vs Instituciones	87
3.7. Exploración de las 5 variables más Influyentes	88
<b>4. Conclusiones y Discusiones</b>	<b>94</b>
4.1. Conclusiones	94
4.2. Discusiones	95
<b>5. Trabajos Futuros</b>	<b>96</b>
<b>6. Referencias Bibliográficas</b>	<b>97</b>
<b>7. ANEXO</b>	<b>100</b>

## Listado de Ilustraciones

Ilustración 1. Minería de datos, Pasos del proceso de descubrimiento de Conocimiento – KKD .....	22
Ilustración 2. Diagrama de dispersión biespacial .....	24
Ilustración 3. Aspecto Visual de un árbol de decisión .....	25
Ilustración 4: DER de Características de los miembros del hogar.....	32
Ilustración 5: DER de Características de la Vivienda.....	33
Ilustración 6: Proceso de ETL para la creación de la tabla analítica .....	34
Ilustración 7: Distribución de opiniones validas según Departamentos del año 2014 .....	35
Ilustración 8: Distribución de Confianza con las instituciones del Gobierno 2014 .....	37
Ilustración 9: Proporción de “Nada confía” del Top 4 de instituciones del 2007 al 2004 .....	38
Ilustración 10: Afinidad por Departamentos según la percepción de confianza del 2014 .....	40
Ilustración 11: Percepción del top 4 Instituciones en segmentos similares por departamentos.....	41
Ilustración 12: Distribución de Top 4 Instituciones, del 2007 al 2014 .....	42
Ilustración 13: Evolución del coeficiente de contingencia sobre las variables más influyentes en la Confianza sobre el Congreso de la República, del 2007 al 2014. ....	45
Ilustración 14: Primera variable más Influyente en la Confianza en el congreso de la República .....	48
Ilustración 15: Segunda Variable más Influyente en la Confianza en el Congreso de la República ....	49
Ilustración 16: Tercera Variable más Influyente en la Confianza en el Congreso de la República .....	50
Ilustración 17: Cuarta Variable más Influyente en la Confianza en el Congreso de la República .....	51
Ilustración 18: Quinta Variable más Influyente en la Confianza en el Congreso de la República .....	52
Ilustración 19: Nivel de estudios y la Confianza en el Congreso de la República, 2007, 2013 y 2014 .....	55
Ilustración 20: Lengua Materna y la Confianza en el Congreso de la República, 2007, 2013 y 2014 .....	55
Ilustración 21: Dominio Geográfico y la Confianza en el Congreso de la República, 2007, 2013 y 2014 .....	56
Ilustración 22: Material Predominante en las paredes y la Confianza en el Congreso de la República, 2007, 2013 y 2014.....	56
Ilustración 23: Combustible que usan en el Hogar para cocinar y la Confianza en el Congreso de la República, 2007, 2013 y 2014.....	57
Ilustración 24: Evolución del coeficiente de contingencia sobre las variables más influyentes en la confianza sobre el Gobierno Regional, del 2007 al 2014. ....	58
Ilustración 25: Primera variable más Influyente en la Confianza en el Gobierno Regional .....	60
Ilustración 26: Segunda variable más Influyente en la Confianza en el Gobierno Regional .....	61
Ilustración 27: Tercera variable más Influyente en la Confianza en el Gobierno Regional .....	62
Ilustración 28: Cuarta variable más Influyente en la Confianza en el Gobierno Regional .....	63
Ilustración 29: Quinta variable más Influyente en la Confianza en el Gobierno Regional .....	64
Ilustración 30: Nivel de estudios y la Confianza del Gobierno Regional, 2007, 2013 y 2014.....	66

Ilustración 31: Lengua Materna y la Confianza del Gobierno Regional, 2007, 2013 y 2014 .....	66
Ilustración 32: Dominio Geográfico y la Confianza del Gobierno Regional, 2007, 2013 y 2014 .....	67
Ilustración 33: El material predominante en los Pisos y la Confianza del Gobierno Regional, 2007, 2013 y 2014 .....	67
Ilustración 34: Evolución del coeficiente de contingencia sobre las variables más influyentes en la confianza sobre la Policía Nacional del Perú, del 2007 al 2014. ....	68
Ilustración 35: Primera variable más Influyente en la Confianza en la Policía Nacional .....	70
Ilustración 36: Segunda variable más Influyente en la Confianza en la Policía Nacional .....	71
Ilustración 37: Tercera variable más Influyente en la Confianza en la Policía Nacional .....	72
Ilustración 38: Cuarta variable más Influyente en la Confianza en la Policía Nacional .....	73
Ilustración 39: Quinta variable más Influyente en la Confianza en la Policía Nacional .....	74
Ilustración 40: Nivel de estudios y la Confianza con la Policía Nacional, 2007, 2013 y 2014 .....	75
Ilustración 41: Lengua Materna y la Confianza con la Policía Nacional, 2007, 2013 y 2014 .....	75
Ilustración 42: Dominio Geográfico y la Confianza con la Policía Nacional, 2007, 2013 y 2014.....	76
Ilustración 43: Material Predominante en las Paredes y la Confianza del Policía Nacional, 2007, 2013 y 2014 .....	76
Ilustración 44: Evolución del coeficiente de contingencia sobre las variables más influyentes en la confianza sobre el Poder Judicial, del 2007 al 2014. ....	77
Ilustración 45: Primera variable más Influyente en la Confianza en el Poder Judicial .....	79
Ilustración 46: Segunda variable más Influyente en la Confianza en el Poder Judicial.....	80
Ilustración 47: Tercera variable más Influyente en la Confianza en el Poder Judicial.....	81
Ilustración 48: Cuarta variable más Influyente en la Confianza en el Poder Judicial .....	82
Ilustración 49: Quinta variable más Influyente en la Confianza en el Poder Judicial.....	83
Ilustración 50: Nivel de estudios y la Confianza en el Poder Judicial, 2007, 2013 y 2014 .....	85
Ilustración 51: Lengua materna y la Confianza en el Poder Judicial, 2007, 2013 y 2014 .....	85
Ilustración 52: Dominio Geográfico y la Confianza en el Poder Judicial, 2007, 2013 y 2014.....	86
Ilustración 53: Material Piso y la Confianza en el Poder Judicial, 2007, 2013 y 2014.....	86
Ilustración 54: Comparativa de Comportamiento de las tres principales variables y las Instituciones 2014.....	87
Ilustración 55: Evolución Nivel de estudios .....	88
Ilustración 56: Lengua Materna .....	89
Ilustración 57: Evolución Dominio Geográfico.....	89
Ilustración 58: Material de Paredes .....	90
Ilustración 59: Combustible usado en el Hogar.....	90
Ilustración 60: Acceso a servicios de internet, TV, Cable. ....	91
Ilustración 61: Material de pisos de la Vivienda .....	91
Ilustración 62: Variables top 5 más influyentes de la base del 2014 con el congreso de la República 92	

## Listado de Tablas

Tabla 1: Ventajas y limitaciones que incluyen cada uno de los métodos .....	27
Tabla 2: Módulos seleccionados sobre la encuesta nacional de hogares ENAHO 2014 .....	30
Tabla 3: Distribución de la Cantidad de Viviendas y Hogares Encuestadas del 2007 al 2014, procesado según los datos disponibles en el repositorio del INEI.....	31
Tabla 4: Distribución de Confianza con las instituciones del Gobierno 2014 .....	36
Tabla 5: Proporciones de la Percepción de la confianza del Congreso de la República, por departamentos.....	39
Tabla 6: Resumen de porcentajes del Top 4 de las instituciones Seleccionadas .....	43
Tabla 7: Resumen de las variables más influyentes en la confianza sobre las 4 Instituciones.....	44
Tabla 8: Resumen de las variables más influyentes en la confianza sobre el Congreso de la República .....	45
Tabla 9: Resumen de las variables más influyentes en la confianza sobre el Gobierno Regional .....	58
Tabla 10: Resumen de las variables más influyentes en la confianza sobre la Policía Nacional del Perú .....	68
Tabla 11: Resumen de las variables más influyentes en la confianza sobre el Poder Judicial .....	77
Tabla 12: Módulos sobre la encuesta nacional de hogares ENAHO 2014 .....	100



## Resumen

Este trabajo consiste en un análisis sobre la percepción de la confianza en las instituciones públicas del gobierno, usando técnicas de Minería de datos, se busca obtener los 5 factores más influyentes en el año 2014, con respecto a la percepción de confianza en las instituciones del gobierno del Perú, y cómo es su comportamiento desde el 2007 al 2014.

Para conseguir este objetivo se usan datos públicos de encuestas, que están disponibles en el repositorio digital del Instituto Nacional de Estadísticas e Informática de Perú (INEI), que, siendo órgano rector de las estadísticas del Perú, ejecuta todos los años un programa de Encuestas Nacional de Hogares (ENAHOG), el cual cuenta con 29 categorías de estudios, siendo algunas de ellas: Educación, Empleabilidad, Características del hogar, Gobernabilidad y Democracia.

El trabajo comenzó con la búsqueda de datos en los repositorios del INEI, continuó con la carga en una base de datos, luego se efectuó la limpieza e integración usando SQL, hasta la construcción de una tabla Analítica. Sobre estos datos se aplicaron métodos descriptivos para la selección de variables potenciales. Luego ejecutando el algoritmo de árbol de decisión, se obtuvieron las 5 variables más influyentes, con el Análisis Factorial de correspondencias se vio las atracciones entre las modalidades de estas variables y la confianza.

Los resultados del trabajo muestran que el Congreso de la República es la institución con mayor desconfianza, seguida por el Gobierno regional, La Policía Nacional y el Poder Judicial. Estas instituciones fueron las que se seleccionaron para realizar el análisis de influencia, no combinando estas ya que las instituciones son autónomas y se sospecha que las influencias son independientes.

Las variables más influyentes son: *[Nivel de estudios]*, *[Lengua Materna]* y *[Dominio Geográfico]* y estas son las mismas para las instituciones estudiadas, las variables restantes son *[El material predominante en las paredes]* y *[El material predominante en los Pisos]*, *[Combustible que usan para cocinar]* y *[No cuenta con teléfono fijo, celular, tv, cable, internet]*. El trabajo también muestra la evolución temporal de influencia, comportamiento de las variables.

# 1. Introducción

Actualmente la enorme cantidad de base de datos en todas las áreas de aplicación humana, demanda nuevas y poderosas técnicas de transformación de los datos en conocimientos útiles, entre dichas técnicas podemos nombrar las pertenecientes al aprendizaje automático, al análisis estadístico de datos, a la visualización de datos y a las redes neuronales. La minería de datos se refiere a la aplicación de estas técnicas para encontrar patrones en los datos. [1].

En la actualidad, los gobiernos de prácticamente todos los países procuran no sólo proteger a sus ciudadanos, sino también velar para que hasta los más pobres gocen de servicios básicos. Por lo general estos servicios incluyen educación, atención básica de la salud y abastecimiento de agua potable. A veces se extienden mucho más y abarcan pensiones de vejez y ayuda al discapacitado.

Los gobiernos establecen la prestación de estos servicios por diversas instituciones estatales, como la policía, departamentos de obras públicas, ministerios de educación, servicios de salud pública, autoridades del agua, etc. Los organismos de ayuda extranjera emplean estas mismas instituciones para encauzar proyectos destinados a beneficiar a los pobres. Sin embargo, desde la perspectiva de los pobres nos encontramos frente a una crisis institucional. Si bien existen casos aislados de excelencia, por lo general los pobres opinan que las instituciones formales carecen de eficacia, son inaccesibles y les anulan su propio potencial. Son temas reiterados en los informes: la desconfianza, corrupción, humillación, intimidación, impotencia, desesperanza y, a menudo, la ira.

La confianza en las instituciones, que constituyen el principal mecanismo para solucionar los conflictos y delimitar las normas que estructuran el comportamiento en sociedad, es un factor sustancial que impacta sobre la estabilidad de la democracia. La percepción que los ciudadanos tienen de las mismas reviste entonces singular importancia, ya que como afirma Miller (1974: 951), “un sistema político no puede sobrevivir por largo tiempo sin el apoyo de la mayoría de sus ciudadanos”.

En este trabajo se usan las técnicas de minería de datos para descubrir patrones de relaciones entre factores que afectan a la percepción de la confianza en las instituciones del gobierno, y para analizar su comportamiento durante los años 2007 al 2014.

Los datos que se usaron para este estudio fueron extraídos de la encuesta Nacional de Hogares en sus siglas (ENAHOG) [2], sobre resultados de opiniones referidos a confianza en las instituciones públicas, y los datos sobre características de la población referidos a características de la vivienda y del hogar, características de los miembros del hogar, educación, salud, empleo e ingresos, Instituciones benéficas, mantenimiento de la vivienda, transporte y comunicaciones, etc. Esta encuesta es ejecutada todos los años por el Instituto Nacional de Estadística e Informática en sus siglas (INEI) que cumple funciones similares al INDEC (Argentina).

El objetivo de este trabajo es realizar un análisis sobre los principales factores que influyen en la percepción de la confianza en las 4 instituciones del gobierno peruano con mayor proporción de desconfianza y evaluar su evolución a lo largo del tiempo

En el análisis de selección de variables se considera las 5 principales más influyentes, determinadas según los datos del 2014, mientras que el análisis temporal se realiza del 2007 al 2014.

## **1.1. Planteamiento del Problema**

Las instituciones públicas del gobierno son el único medio por el cual el gobierno se relaciona directamente con la población, un alto porcentaje de desconfianza, hace que la gobernabilidad se vea afectada, en un país con políticas de gobernabilidad democrática.

¿Que deberían hacer los gobiernos regionales para que el pueblo sienta una confianza total en las instituciones públicas?, ¿será el nivel de estudios que influye y el que permita una mejor relación entre el gobierno y el pueblo?, ¿las instituciones públicas realizan campañas para mejorar su relación con el pueblo?, ¿será que existe funcionarios públicos que brindan los servicios en zonas donde la lengua de la mayoría no es el castellano y que estos funcionarios solamente hablan castellano?, ¿cómo podemos segmentar a estas personas que sienten desconfianza en el gobierno?, ¿será posible que las técnicas de minería de datos permitan determinar cuál de los factores siguientes: nivel de estudios, Empleabilidad, Salud, Características y percepción del hogar, entre otras, sea el más influyente en la percepción de confianza en las instituciones del gobierno?

Nos planteamos preguntas de investigación.

- **¿Cuáles son los factores más influyentes en la percepción negativa sobre las instituciones del gobierno en el Perú en el año 2014?**
- **¿Cómo es el comportamiento de los factores más influyentes con la percepción de la confianza en las instituciones del gobierno Peruano en el lapso 2007 - 2014?**

## **1.2. Justificación del Problema**

La confianza en las instituciones del gobierno permitirá una buena relación con el pueblo, por ello el gobierno deberá impulsar programas sociales para mejorar la educación, empleo, el gobierno central y regional serán los beneficiados directamente de este trabajo, el que permitirá tomar decisiones estratégicas para mejorar la relación con la población.

Una crisis de la confianza institucional, puede amenazar las bases del Estado democrático de Derecho, así como su grado de influencia y forma de actuación en la construcción de la confianza de los ciudadanos hacia las instituciones públicas.

Este trabajo tiene mucha transcendencia social y al utilizar métodos de minería de datos se espera que pueda servir como modelo para futuras investigaciones de esta categoría.

## **1.3. Antecedentes del Problema**

Hace más de una década, al término del mandato presidencial de Alberto Fujimori, el Perú entró en una estabilidad y crecimiento en la economía, junto con los gobiernos de Alejandro Toledo y Alan García, pero aun es un desafío alcanzar un estado eficiente e inclusivo. Por lo que es necesario promover los programas sociales, mejorar la educación, fortalecer la gobernabilidad democrática, la gestión pública, la seguridad ciudadana, la lucha contra la corrupción para lograr una democracia plena.

Marisol Espinoza, Vicepresidenta de la República del Perú, indicó que es necesario “abrir las puertas del diálogo”. “La democracia se construye, parte de la democracia es entender al otro, entender que el otro también es un ciudadano que necesita ser escuchado. Cuando seamos capaces desde el Estado de entender que hay que asumir

un papel activo en el tema de los conflictos, creo que vamos a dar un gran paso adelante en la resolución de la conflictividad en el país”. [3]. Publicado el 28 de noviembre del 2012

Según resultados sobre incremento de la democracia 2013 medidos respecto al promedio de 1995-2013 en puntos porcentuales, el Perú está ubicado en el último puesto con un incremento de 2% respecto al promedio medido el 1995-2013, esto medido según el latino-barómetro [4]

Según el diario peruano la república en una nota periodística, indica que “somos los menos entusiastas con la afirmación de que la democracia es el mejor sistema de gobierno, junto con México y El Salvador; pensamos que nuestra democracia tiene “grandes problemas”, junto con Honduras, Paraguay y Brasil; somos los más insatisfechos con el funcionamiento de la democracia,” [5]

El actual presidente Ollanta Humala Taso (periodo de gobierno 2012 al 2016) en el discurso titulado “Gobierno peruano lucha por el desarrollo del país para recuperar confianza de la población” dijo que la gran transformación es dar oportunidades a todos los peruanos para reducir la brecha de la desigualdad. Por ello, explicó que su gobierno impulsa la educación y la ejecución de infraestructura en todo el país<sup>1</sup>. Publicado el 10 junio, 2013.

#### **1.4. Alcances y Limitaciones**

Este trabajo tiene como alcance solo los datos de la Encuesta Nacional de Hogares (ENAHOG) del 2007 al 2014, en sus Módulos de Características de Hogar, Educación, Empleabilidad, Salud, Gobernabilidad (Confianza de las instituciones del Gobierno), y otros. Como se ve en la tabla 2, donde se seleccionan las variables que proporcionan un análisis positivo.

---

<sup>1</sup> <http://www.presidencia.gob.pe/presidente-humala-gobierno-peruano-lucha-por-el-desarrollo-del-pais-para-recuperar-confianza-de-la-poblacion>

## **1.5. Objetivo de Investigación**

### **Objetivo General**

- Determinar factores que influyen en la percepción de la confianza en las instituciones públicas del gobierno en el Perú del 2014, y analizar estos temporalmente durante los años 2007 al 2014.

### **Objetivos Particulares**

- Usar el proceso de descubrimiento de conocimiento utilizado en minería de datos para resolver el problema planteado.
- Realizar un análisis de comportamiento de influencia de los factores, sobre la confianza en las 4 instituciones públicas con mayor porcentaje de desconfianza.
- Realizar un análisis de temporalidad sobre los factores más influyentes, en el comportamiento de la confianza en las instituciones desde el 2007 al 2014.

## 2. Marco Teórico

### 2.1. Estado del arte

Actualmente existen muchos trabajos relacionados a la confianza en las instituciones de los gobiernos públicos, estas son desarrolladas por organismos internacionales como CEPAL, PNUD, también por algunos países como México y El Salvador. También se encuentran trabajos de tesis. De los mismos se detallan algunos que están más relacionados con este trabajo, si bien los resultados que proporcionan son obtenidos mediante técnicas de análisis tradicionales, usados en la estadística descriptiva, no incorporando técnicas avanzadas de exploración como las usadas en la minería de datos.

Según el Boletín de PNUD& Instituto de Iberoamérica año 2007, se realizó un estudio de análisis sobre el impacto del crecimiento económico en la confianza en las instituciones de gobierno y ciudadanía y de los parlamentarios latinoamericanos en 12 casos nacionales. Para este trabajo se usaron bases de encuestas para ciudadanos del Proyecto de Opinión Pública de América Latina (LAPOP) de la Universidad de Vanderbilt respecto a sus opiniones sobre la confianza en las instituciones y la encuesta de las bases de datos Elites Parlamentarias de América Latina (PELA) desarrollada desde el Instituto de Iberoamérica de la Universidad de Salamanca para conocer la opinión de los legisladores. La hipótesis general es que a mayor crecimiento económico mayor será la confianza en las instituciones tanto desde los ciudadanos como de las élites latinoamericanas. Como resultados del análisis de Confianza de los ciudadanos en las instituciones de gobierno 2004-2011, surge que Perú y Paraguay conforman los casos con peores evaluaciones de sus instituciones, por debajo del 20%. [6]

En un trabajo realizado en Argentina año 2007, categorizado como análisis político, que se presentó para el VII Congreso nacional de ciencia política, con el tema denominado “CONFIANZA PÚBLICA, PARTICIPACIÓN CIUDADANA Y APEGO A LA LEY EN LA ARGENTINA POSTDEVALUACIÓN”. “El proceso de recuperación económica iniciado por la Argentina en los últimos años, ha dado lugar a una sostenida recomposición de la confianza en las instituciones públicas, aunque esto todavía ocurre en un contexto de generalizado descreimiento en el funcionamiento de las mismas. Un punto sobre el cual cabe llamar la atención es que el aumento de la confianza pública no pareció ir acompañado por una mayor participación ciudadana, que se mantuvo en

niveles muy bajos e incluso tendió a menguar en las clases medias. Es significativo, al respecto, lo observado en relación a la participación en las actividades de voluntariado o filantropía, que se redujeron considerablemente. Se advierte, en consecuencia, que las mejoras económicas y sociales recientes no se han visto reflejadas en una mayor responsabilidad y conciencia ciudadana, con el consiguiente riesgo de relegar de la agenda de discusión pública la deuda aún pendiente en materia de justicia, cohesión e integración social”. [7]

En el Salvador, marzo del 2013, se presentó el informe denominado “La percepción de la seguridad y la Confianza en las instituciones públicas”, como línea base del plan de acción Asocio para el crecimiento. Conducido por el instituto Universitario de Opinión (IUDOP), de la universidad Centroamericana “José Simeón Cañas” (UCA). a solicitud de la Agencia de los Estados Unidos para el Desarrollo Internacional (USAID) bajo los objetivos del Plan de Acción Conjunto El Salvador-Estados Unidos Asocio para el Crecimiento. El propósito general del estudio fue conocer la percepción de los salvadoreños y salvadoreñas sobre la situación de la seguridad y la confianza hacia las instituciones públicas, a fin de establecer la Línea de Base de los indicadores relativos a las percepciones públicas sobre el crimen y la inseguridad, que permitan evaluar posteriormente el progreso de las Metas 1, 3, 4, 6 y 7 del Plan Asocio para el Crecimiento. Para ello, se cursaron dos encuestas de opinión entre el 9 y 18 de septiembre de 2012. La primera de ellas es una encuesta realizada a una muestra nacional de 2.413 adultos, la cual es representativa del conjunto de la población salvadoreña de 18 años a más; este estudio cuenta con un 95 por ciento de confiabilidad y un error muestral de más o menos 1.99 por ciento (+/-0.019). Paralelamente a este sondeo, fue realizada una encuesta a una muestra de 521 Micro y Pequeños empresarios, la cual contó con un error muestral de más o menos 4.9 por ciento (+/-0.049).

Se crearon indicadores metas para la confianza en escala 0 a 100, en el que 0 representan ausencia completa de confianza, y 100 expresa una completa confianza. Los resultados son gráficos descriptivos de 2 entradas, con el objetivo de relacionar la confianza con variables demográficas. En este estudio se encontraron variaciones en función de la zona de residencia del entrevistado, la edad, el grado de escolaridad, la estratificación social, el ingreso y la exposición a noticias. En esta oportunidad, el sexo no marcó diferencias estadísticamente significativas. A su vez, los que han sido afectados por un delito recientemente y los que se sienten más inseguros expresan menor confianza hacia las instituciones públicas.



Un cruce entre la zona de residencia y el Índice de confianza hacia la institucionalidad pública muestra que nuevamente los residentes de las zonas rurales expresan promedios más altos de confianza (51.7), en una escala de 0 a 100, respecto a los habitantes de las urbes (48.5), quienes se ubican por debajo del promedio de la muestra (50.1). Por otra parte, las personas más jóvenes parecen tener mayor confianza hacia las instituciones gubernamentales que el resto de grupos etáreos. El grupo de edad que expresa los más bajos niveles de confianza pública son los que tienen entre 41 a 55 años. Esto es interesante porque tradicionalmente se ha caracterizado a los jóvenes como el grupo con mayores niveles de Anomia, distantes y con bajo interés en el ámbito institucional. [8]

En el año 2014 el gobierno de México, realizó una encuesta telefónica sobre la confianza en las instituciones, la encuesta fue realizada por medio de encuestas telefónicas (CATI) a 623 personas con teléfonos fijos del hogar, con un error de marco muestral del +/-3.9% y nivel de confianza del 95%, con resultados comparativos de los últimos dos años anteriores. [9]

- Casi todas las instituciones incluidas en el estudio no muestran un cambio significativo respecto al año anterior. Sin embargo, la mayoría de las instituciones muestran una ligera tendencia a la baja.
- Las instituciones en las que señalan confiar más los entrevistados son: la familia, los amigos, las universidades públicas, la Cruz Roja y la Marina.
- En las que menos confían son la policía estatal, la televisión, la Cámara de Diputados, los sindicatos y los partidos políticos.
- Entre la medición del 2013 y la de este año, las instituciones que tuvieron una caída mayor al 5% de los que señalaron que confían “mucho o “algo” en ellas fueron: el ejército, la marina, el Presidente de la República, los gobernadores, los partidos políticos, los hospitales, las pequeñas y las grandes empresas. Únicamente los policías estatales y municipales tuvieron un aumento mayor al 5%. El resto de las instituciones tiene cambios menores.
- 8 de cada 10 (77%) confía mucho en su familia. En el caso de los amigos, este porcentaje se reduce a 1 de cada 2 (52%) y sólo 1 de cada 3 (35%) confía mucho en sus vecinos.
- Las fuerzas castrenses pese a tener una ligera caída aún generan confianza entre los mexicanos. 68% confía “mucho” o “algo” en la Marina y el 61% es el nivel de

confianza en el ejército. Existe una leve mejoría en la confianza que la población tiene en las policías estatal y municipal.

- No se registran cambios en la confianza que la población que tiene en el Instituto Nacional electoral (INE), pese a haber atravesado un proceso de transformación al dejar de ser el Instituto Federal Electoral (IFE).
- Más de la mitad de los entrevistados (56%) confía mucho o algo en la Comisión Nacional de los Derechos Humanos - CNDH.
- Entre agosto del 2012 y el mismo mes del 2013, la confianza que la población tenía en los partidos políticos aumentó considerablemente, al pasar del 16% al 29% que señalaba que confiaba “mucho” o algo” en estas instituciones; sin embargo, este porcentaje entre el 2013 y el 2014 cayó al pasar del 29% al 22% que tiene esta opinión.

## **2.2. Confianza Ciudadana**

Puede entenderse como confianza social la creencia en que una persona o grupo será capaz y deseará actuar de manera adecuada en una determinada situación, lo cual supone un cierto grado de regularidad y predictibilidad de las acciones que facilitan el funcionamiento de la sociedad. Como tal, es una predisposición hacia el otro que se aprende, y desde la perspectiva del capital social, incluye un juicio moral acerca del grado de confianza que puede tenerse respecto de las personas desconocidas. Hay que distinguir entre “confianza generalizada” o confianza social respecto de personas o instituciones desconocidas sobre las cuales se dispone de insuficiente información o experiencias para hacer un juicio, y la “confianza particularizada” o interpersonal, referida a la confianza en personas que se conoce a partir de “lazos débiles”. La desconfianza social, se origina fundamentalmente en la experiencia personal o ajena de que los otros no hacen lo esperado, que no cumplen con las normas básicas. Un elemento que socava la confianza en las instituciones es el uso discrecional o excesivo del poder y la influencia (CEPAL, 2007) [7].

## **2.3. Las Instituciones Públicas del Perú**

En el portal del Estado peruano, se pueden encontrar el listado de todas las instituciones, y estas se pueden dividir en grupos como son: Ejecutivo, Legislativo, Judicial, Organismos Autónomos, Gobiernos Locales y Gobiernos Regionales. El listado está disponible en la siguiente url: <http://www.peru.gob.pe>, en este trabajo solo se detalla algunas de ellas.

### **2.3.1. Oficina Nacional de Procesos Electorales Onpe<sup>2</sup>**

La Oficina Nacional de Procesos Electorales - ONPE es un organismo electoral constitucional autónomo que forma parte de la estructura del Estado. Es la autoridad máxima en la organización y ejecución de los procesos electorales, de referéndum y otros tipos de consulta popular a su cargo. Su finalidad es velar por que se obtenga la fiel y libre expresión de la voluntad popular, manifestada a través de los procesos electorales a su cargo. Con relación a las organizaciones políticas, se encarga de la verificación de firmas de adherentes de los partidos políticos en proceso de inscripción; la verificación y control externos de la actividad económico-financiera, así como brindar asistencia técnico-electoral en los procesos de democracia interna.

Asimismo, otorga apoyo y asistencia técnico-electoral en la elección de autoridades de centros poblados, de universidades, de representantes de la sociedad civil en los consejos de coordinación regional y local y de autoridades y representantes de todas las organizaciones de la sociedad y del Estado que requieran su apoyo en los procesos electorales.

Promueve la cultura electoral democrática a través de sesiones de educación electoral y, como parte de ello, específicamente, el respeto de los derechos políticos de la mujer en cumplimiento a la política de igualdad de oportunidades entre hombres y mujeres y otras que las leyes señalan.

Desarrolla y promueve la investigación electoral, mediante la producción de documentos de información sistematizada de los procesos electorales y de consulta realizados, así como la publicación y difusión de investigaciones orientadas al tema.

---

<sup>2</sup> <http://www.web.onpe.gob.pe/quienes-somos.html>

### **2.3.2. Registro Nacional de Identificación y Estado Civil RENIEC<sup>3</sup>**

Organismo autónomo encargado de la identificación de los peruanos, otorga el documento nacional de identidad, registra hechos vitales: nacimientos, matrimonios, defunciones, divorcios y otros que modifican el estado civil. En procesos electorales su participación consiste en proporcionar el Padrón Electoral que será utilizado el día de las elecciones, fue creada constitucionalmente mediante la Ley N° 26497, con fecha 12 de julio de 1995,

En este contexto las políticas y esfuerzos institucionales están orientados a:

- Documentar a los peruanos indocumentados adultos, en el marco de lo establecido en el Plan Nacional de Restitución de la Identidad, previniendo la indocumentación y restituyéndola de ser el caso. En este contexto se ha logrado documentar gratuitamente a más de 5 millones de ciudadanos peruanos (menores y mayores de edad, personas con discapacidad, adultos mayores) entre los años 2002-2009.
- Documentar a los menores de edad con el DNI, estableciendo alianzas estratégicas con el Ministerio de Educación y entidades educativas públicas y privadas
- Fortalecer el proceso de registro y actualización del registro único de identificación de personas mediante la implementación de soluciones tecnológicas como el sistema automatizado de identificación por impresiones dactilares (AFIS) que han permitido validar la unicidad e integridad de sus registros.
- Garantizar la identificación y el registro de hechos vitales de los peruanos, mediante la utilización de tecnologías y procedimientos de alta calidad; con el compromiso de cumplir los requisitos y mejorar continuamente la eficacia del Sistema de Gestión de Calidad.

---

<sup>3</sup> <http://www.reniec.gob.pe/portal/institucional.htm#>

### **2.3.3. La Municipalidad Provincial y Distritales**

- Cada provincia, distrito y centro poblado tienen su gobierno local o Municipalidad.
- Las funciones de las municipalidades son las siguientes:
  - Decidir cómo se usarán los terrenos y los predios de su localidad. Por ejemplo, el tamaño de los lotes de terreno depende de los permisos que otorguen las municipalidades; asimismo, el uso que se dará a las casas, si son sólo para residencia o para uso comercial.
  - Encargarse de los servicios públicos básicos. Por ejemplo, la limpieza de las calles, hacer campañas de salud y recolectar la basura.
  - Protección y conservación del medio ambiente. Por ejemplo, combatir la contaminación ambiental y cuidar el patrimonio cultural, histórico y paisajístico de su localidad.
  - Promover el desarrollo económico de su localidad.
  - Promover la participación vecinal, sobretodo en el uso y control de los recursos.
  - Organizar servicios sociales para proteger a la población de menores recursos.
  - Luchar contra la comercialización y consumo de drogas.

### **2.3.4. Policía Nacional del Perú<sup>4</sup>**

La Policía Nacional del Perú es una institución del Estado que tiene por misión garantizar, mantener y restablecer el orden interno, prestar protección y ayuda a las personas y a la comunidad, garantizar el cumplimiento de las leyes y la seguridad del patrimonio público y privado, prevenir, investigar y combatir la delincuencia; vigilar y controlar las fronteras; con el propósito de defender a la sociedad y a las personas, a fin de permitir su pleno desarrollo, en el marco de una cultura de paz y de respeto a los derechos humanos.

---

<sup>4</sup> <https://www.pnp.gob.pe/nosotros.html>

### **2.3.5. Fuerzas Armadas**

Es el organismo que tiene como misión defender al Perú de sus amenazas y proteger sus intereses, asumir el control del orden interno, participar en el desarrollo económico y social del país y en la defensa civil de acuerdo a la ley mediante el empleo del poder aeroespacial a fin de contribuir a garantizar su independencia, soberanía e integridad territorial.

### **2.3.6. Congreso de La República<sup>5</sup>**

El Congreso de la República representa la opinión pluralista de la nación. Todas las sangres, credos, ideologías, profesiones y oficios están representados por el voto democrático en la proporción que la ciudadanía decide.

Los congresistas han sido elegidos para servir a la nación, legislar, fiscalizar y representar nuestros intereses y aspiraciones, por eso gozan del privilegio de la inmunidad y no pueden ser responsabilizados por sus opiniones y votos, ni ser procesados sin autorización del Congreso. El ciudadano debe ver en el congresista un mediador entre él y el Estado.

El trabajo del congresista es a tiempo completo, salvo su contribución a la formación de la juventud en la docencia universitaria. Los actos del congresista y del Congreso son públicos. La ciudadanía los conoce a través de los medios de comunicación.

### **2.3.7. Poder Judicial del Perú<sup>6</sup>**

Según el portal del estado peruano indica que la misión de esta institución es: Administrar Justicia a través de sus órganos jurisdiccionales, con arreglo a la Constitución y a las leyes, garantizando la seguridad jurídica y la tutela jurisdiccional, para contribuir al estado de derecho, al mantenimiento de la paz social y al desarrollo nacional

---

<sup>5</sup> <http://www.congreso.gob.pe/sobreelcongreso>

<sup>6</sup> [http://www.peru.gob.pe/directorio/pep\\_directorio\\_detalle\\_institucion.asp?cod\\_institucion=10051](http://www.peru.gob.pe/directorio/pep_directorio_detalle_institucion.asp?cod_institucion=10051)

## 2.4. Encuesta de Hogares

Según la ficha técnica de este estudio publicado en el repositorio del INEI (<http://inei.inei.gob.pe/microdatos/>), menciona que la Encuesta Nacional de Hogares (ENAHOG) es la investigación que permite al Instituto Nacional de Estadística e Informática (INEI) desde el año 1995, efectuar el seguimiento de los indicadores sobre las condiciones de vida, en el 2014 está formado por 29 módulos de investigación que se detallan en el Anexo.

A inicios del año 2007, con el auspicio del Banco Mundial (BM), el INEI convocó a un Comité Asesor especializado conformado por representantes de organismos internacionales, de organismos gubernamentales nacionales, representantes de la comunidad académica y de centros de investigación. El Comité Asesor de Pobreza ha venido participando cada año en verificar la calidad de la ENAHOG y la medición de la pobreza.

A partir del año 2010, mediante Resolución Suprema N° 097-2010-PCM, publicado en el diario oficial El Peruano el 13 de abril del 2010, el Comité Asesor de Pobreza se constituye en Comisión Consultiva para Estimación de la Pobreza y otros indicadores relacionados en el país; cuyo objetivo es garantizar la calidad, transparencia y confianza de la información en el campo de la medición de la pobreza y otros indicadores relacionados. En el mismo año a iniciativa del Ministerio de Economía y Finanzas se inician las coordinaciones para conformar una Comisión Técnica Interinstitucional que tendría a cargo la Tercera Revisión del Cuestionario de la ENAHOG. Durante los años 2010 y 2011 se realizaron las reuniones del Comité de Revisión y posteriormente se validaron las propuestas de modificación en pruebas de campo. Finalmente en octubre del 2011 se aprobaron los cambios para el cuestionario que rigen a partir del año 2012.

La encuesta se realiza en el ámbito nacional, en el área urbana y rural, en los 24 departamentos del país y en la Provincia Constitucional del Callao.

La población de estudio está definida como el conjunto de todas las viviendas particulares y sus ocupantes residentes en el área urbana y rural del país. Por no ser parte de la población de estudio, se excluye a los miembros de las fuerzas armadas que viven en cuarteles, campamentos, barcos, y otros. También se excluye a las personas que residen en viviendas colectivas (hoteles, hospitales, asilos y claustros religiosos, cárceles, etc.).

El marco muestral para la selección de la muestra lo constituye la información estadística proveniente de los Censos de Población y Vivienda y material cartográfico actualizado para tal fin.

En la muestra no panel se visitan cada año los mismos conglomerados en el mismo mes de la encuesta, pero se seleccionan distintas viviendas.

El nivel de confianza de los resultados muestrales, es del 95%.

La unidad de investigación está constituida por:

- Los integrantes del hogar familiar,
- Los trabajadores del hogar con cama adentro, reciban o no pago por sus servicios,
- Los integrantes de una pensión familiar que tienen como máximo 9 pensionistas, y Las personas que no son miembros del hogar familiar, pero que estuvieron presentes en el hogar los últimos 30 días.

No serán investigados:

- Los integrantes de una pensión familiar que tiene de 10 a más pensionistas, y
- Los trabajadores del hogar con cama afuera.

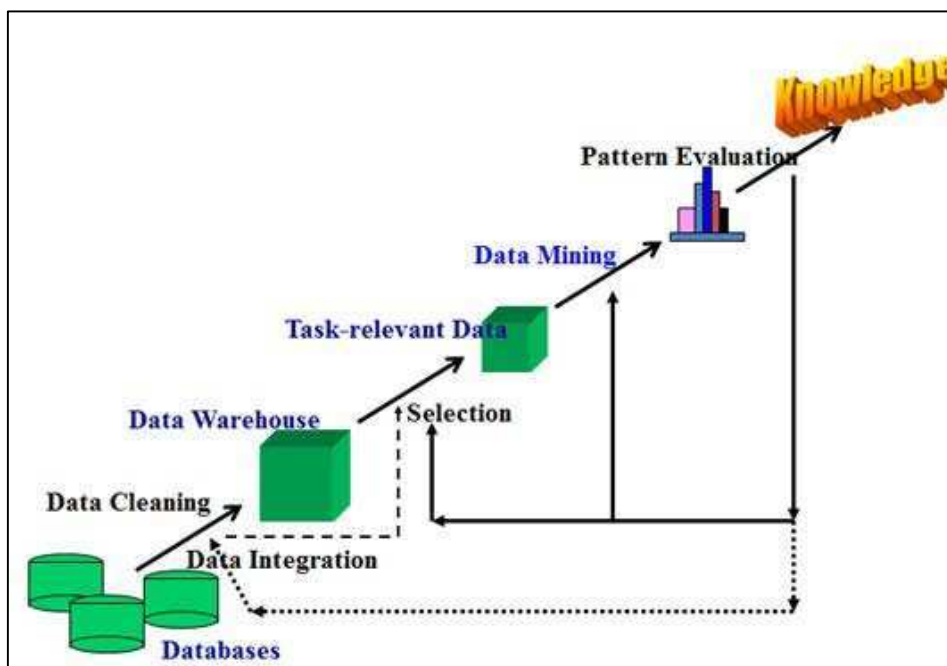
## **2.5. Minería de Datos**

La minería de datos proporciona un nuevo sistema para el manejo de grandes bases de datos, gracias a un campo situado entre la estadística, la tecnología de bases de datos, los patrones de reconocimiento, el aprendizaje de las máquinas, la visualización de datos y los sistemas expertos. Las bases de datos se organizan de forma que puedan ser fácilmente consultadas, actualizadas o utilizadas para extraer información. Un administrador de bases de datos permite realizar estas funciones, con pequeñas bases pero sobre todo con enormes cantidades de datos, difíciles de manejar con otros métodos. La minería de datos no necesita ningún número determinado de variables ni estima que éstas se comporten de una manera concreta mediante funciones de distribución. Así mismo combina gran cantidad de técnicas y métodos, y aplica sus algoritmos, directamente sobre los datos, lo que aporta un manejo y procesamiento de información más eficiente. Se aplican así algoritmos o razonamientos que, originalmente no fueron diseñados para ese tipo de datos, aunque se tiende a modificar la aplicación de forma que la terminología y el campo de trabajo se adecuen a la nueva aplicación [1].



Algunos consideran que Minería de datos es el proceso de, descubrimiento de conocimiento a partir de datos, o KDD. Por otra parte, otros consideran a la minería de datos simplemente como un paso esencial en el proceso de descubrimiento de conocimiento. El descubrimiento de conocimiento como un proceso se representa en la ilustración 1 y se compone de una secuencia iterativa de los pasos siguientes [10]:

1. **Limpieza de datos** (eliminar ruidos e inconsistencia de los datos).
2. **Integración de datos** (múltiples fuentes de datos pueden ser combinados).
3. **Selección de datos** (datos relevantes para la tarea de análisis son extraídos desde las bases de datos).
4. **Transformación de datos** (donde los datos son transforman o consolidan en formatos apropiados para el minado, mediante la realización de operaciones de resumen o de agregación).
5. **Minería de datos** (es el proceso donde son aplicados métodos inteligentes en orden de extraer patrones de datos).
6. **Evaluación de patrones** (identificar los patrones verdaderamente interesantes que representan el conocimiento en base a algunas medidas de Intereses).
7. **Presentación del Conocimiento** (técnicas de visualización y representación del conocimiento, se utilizan para presentar el conocimiento extraído para el usuario).



**Ilustración 1.** Minería de datos, Pasos del proceso de descubrimiento de Conocimiento – KKD

## 2.6. Chi-Cuadrado

El test Chi-cuadrado de independencia fue presentado en 1900 por Karl Pearson (1887-1936), quien ha sido llamado fundador de la ciencia estadística. Un investigador puede estar interesado en saber, respecto de una población, si dos criterios de clasificación están probablemente relacionados o no. Por ejemplo, podemos estar interesados en saber si hay o no asociación entre la religión y el interés político en algún grupo de personas. Si concluimos que dos criterios de clasificación no están relacionados, decimos que son independientes. Dos criterios de clasificación son independientes si la distribución de un criterio no depende de ninguna manera de la distribución del otro. [11]

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \left[ \frac{(f_{oij} - f_{eij})^2}{f_{eij}} \right] \text{ Con } (c-1) (r-1) \text{ grados de libertad.}$$

Donde  $e_{ij} = n_i * n_j / n$

r=Número de filas

c=Número de columnas

### Prueba de chi-cuadrado

El procedimiento Prueba de chi-cuadrado tabula una variable en categorías y calcula un estadístico de chi-cuadrado. Esta prueba de bondad de ajuste compara las frecuencias observadas y esperadas en cada categoría para contrastar que todas las categorías contengan la misma proporción de valores o que cada categoría contenga una proporción de valores especificada por el usuario [12].

### Coeficiente de contingencia

Medida de asociación basada en chi-cuadrado. El valor varía entre 0 y 1. El valor 0 indica que no hay asociación entre las variables de fila y de columna. Los valores cercanos a 1 indican que hay gran relación entre las variables. El valor máximo posible depende del número de filas y columnas de la tabla. [12]

Con el fin de emplear en forma práctica la magnitud del indicador del grado de correlación o asociación entre las variables se da los valores de: 0 (No existe asociación), <0-0,5] (Asociación débil), <0,5-0,8] (Asociación moderada), <0,8-1] (Asociación perfecta).

## 2.7. Análisis Factorial de Correspondencias

El Análisis Factorial de Correspondencias fue ideado por el estadístico francés Benzecri en 1973. Este análisis posibilita profundizar en el análisis de la similaridad de modalidades, de la asociación entre atributos y de las relaciones de atracción y repulsión de sus modalidades en tablas de contingencia construidas comúnmente sobre variables de tipo cualitativo. [13]

El análisis Factorial de Correspondencias, trata de resolver el problema de explicar de forma simple como se atraen o repelen las modalidades de las variables cualitativas observadas. Además del análisis de atracción - repulsión entre modalidades de atributos (variables cualitativas) diferentes, la técnica del análisis Factorial de Correspondencias también permite realizar estudios de proximidad (similaridad/disimilaridad) entre las modalidades de una misma variable; es decir, permite evaluar el parecido de las distribuciones de casos o individuos que presentan modalidades comparadas de una misma variable, proporcionándonos información necesaria para evaluar la homogeneidad o sustituibilidad de las mismas. [13]

Ejemplo:

Se muestra un ejemplo de dispersión biespacial entre las variables Estado civil y Preocupaciones aspecto que más le preocupa, la ilustración 2 muestra filas y columnas juntas, podemos sacar las conclusiones finales, que los Casados prefieren la Armonía Familiar y la Salud frente a los Solteros que prefieren el dinero y la Vida Afectiva.

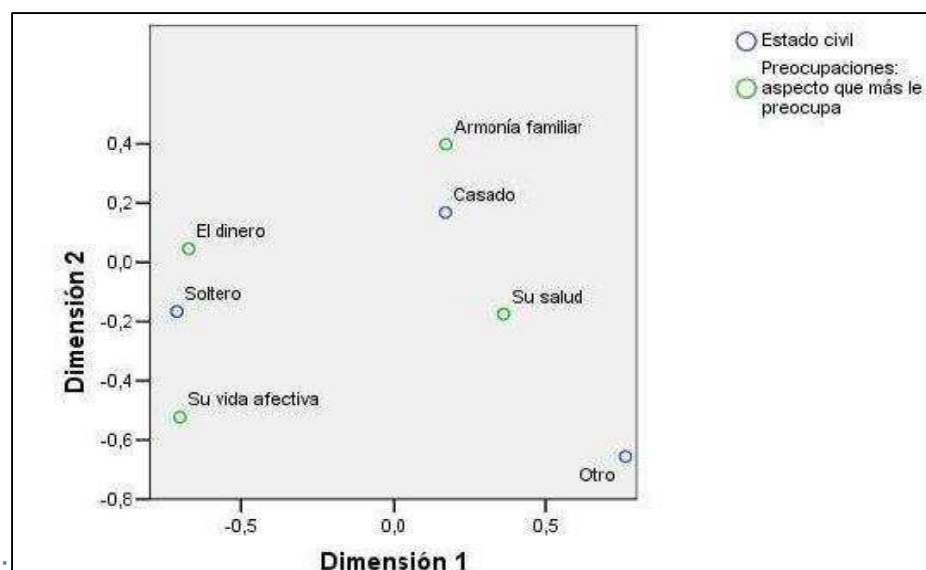


Ilustración 2. Diagrama de dispersión biespacial

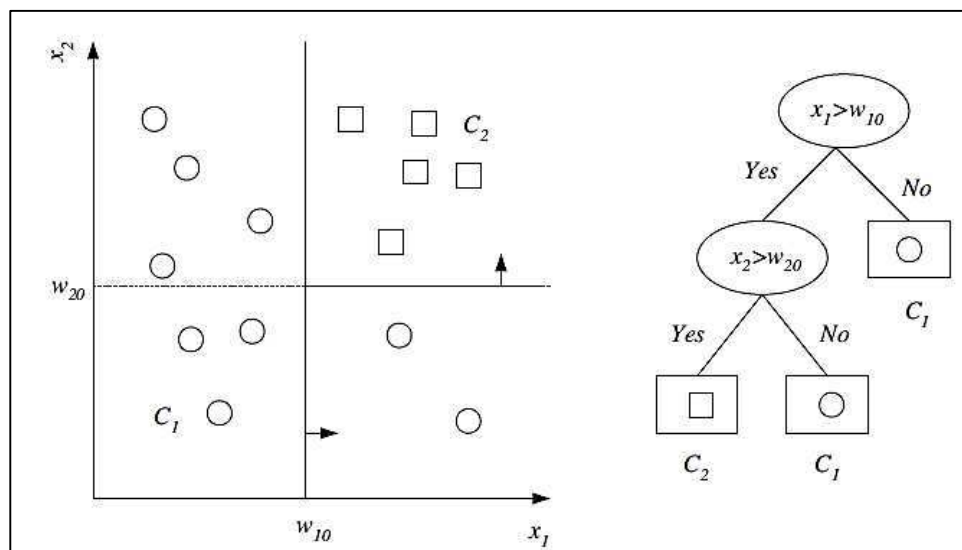
Así pues, los objetivos de esta técnica son:

- Descubrir las relaciones de atracción - repulsión existente entre las distintas modalidades de diferentes variables cualitativas enfrentadas en una tabla de contingencia.
- Descubrir las relaciones de proximidad existentes entre las distintas modalidades de una misma variable cualitativa.
- Visualizar y caracterizar de forma simple las relaciones anteriores en un espacio de dimensión lo más reducida posible.

## 2.8. Árboles de Decisión

Un árbol de decisión es una estructura de datos jerárquica que aplica la estrategia divide y vencerás. Se trata de un método no paramétrico eficiente, que puede ser utilizado tanto para la clasificación y como para la regresión. Discutimos algoritmos que construyen el árbol a partir de una muestra de entrenamiento dado, Como resultado de un árbol se puede obtener un conjunto de reglas simples que son fáciles de entender. [14]

En la ilustración 3 se ve un ejemplo de un conjunto de datos y el árbol de decisión correspondiente. Los óvalos son los nodos de decisión y los rectángulos son nodos hoja. El nodo de decisión univariado se divide a lo largo de un eje, y divisiones sucesivas son ortogonales entre sí. Después de la primera división,  $\{x|x_1 < w_{10}\}$  es puro y no se divide más. [14]



**Ilustración 3.** Aspecto Visual de un árbol de decisión

La mayoría de los algoritmos de construcción de árboles de decisión se basa en la estrategia de Ascenso a la Colina (Hill Climbing). Esta es una técnica utilizada en Inteligencia Artificial para encontrar los máximos o mínimos de una función mediante una búsqueda local. Estos algoritmos empiezan con un árbol vacío, después se va particionado en conjuntos de ejemplos, eligiendo en cada caso aquel atributo que mejor discrimina entre las clases, hasta que se completa el árbol.

## **Métodos de Crecimiento**

Los métodos de crecimiento disponibles son:

**CHAID.** Detección automática de interacciones mediante chi-cuadrado (CHi-square Automatic Interaction Detection). En cada paso, CHAID elige la variable independiente (predictora) que presenta la interacción más fuerte con la variable dependiente. Las categorías de cada predictor se funden si no son significativamente distintas respecto a la variable dependiente. [15]

**CHAID exhaustivo.** Una modificación del CHAID que examina todas las divisiones posibles de cada predictor.

**CRT.** Árboles de clasificación y regresión. CRT divide los datos en segmentos para que sean lo más homogéneos posible respecto a la variable dependiente. Un nodo terminal en el que todos los casos toman el mismo valor en la variable dependiente es un nodo homogéneo y "puro".

**QUEST.** Árbol estadístico rápido, insesgado y eficiente (Quick, Unbiased, Efficient Statistical Tree). Método rápido y que evita el sesgo que presentan otros métodos al favorecer los predictores con muchas categorías. Sólo puede especificarse QUEST si la variable dependiente es nominal.

Tabla 1: Ventajas y limitaciones que incluyen cada uno de los métodos

	CHAID*	CRT	QUEST
Basado en chi-cuadrado**	X		
Variables (predictoras) independientes sustitutas		X	X
Poda de árboles		X	X
División de nodos multinivel	X		
División de nodos binarios		X	X
Variables de influencia	X	X	
Probabilidades previas		X	X
Costes de clasificación errónea	X	X	X
Cálculo rápido	X		X

\*Incluye CHAID exhaustivo.

\*\*QUEST también utiliza una medida de chi-cuadrado para variables independientes nominales.

Fuente: <http://web.udl.es/Biomath/Bioestadistica/SPSS/v17/SPSS%20Decsion%20Trees%2017.0.pdf>

El procedimiento se puede utilizar para:

**Segmentación.** Identifica las personas que pueden ser miembros de un grupo específico.

**Estratificación.** Asigna los casos a una categoría de entre varias, por ejemplo, grupos de alto riesgo, bajo riesgo y riesgo intermedio.

**Predicción.** Crea reglas y las utiliza para predecir eventos futuros, como la verosimilitud de que una persona cause mora en un crédito o el valor de reventa potencial de un vehículo o una casa.

**Reducción de datos y clasificación de variables.** Selecciona un subconjunto útil de predictores a partir de un gran conjunto de variables para utilizarlo en la creación de un modelo paramétrico formal.

**Identificación de interacción.** Identifica las relaciones que pertenecen sólo a Subgrupos específicos y las especifica en un modelo paramétrico formal.

**Fusión de categorías y discretización de variables continuas.** Vuelve a codificar las variables continuas y las categorías de los predictores del grupo, con una pérdida mínima de información.

**Ejemplo.** Un banco desea categorizar a los solicitantes de créditos en función de si representan o no un riesgo crediticio razonable. Basándose en varios factores, incluyendo las valoraciones del crédito conocidas de clientes anteriores, se puede generar un modelo para pronosticar si es probable que los clientes futuros causen mora en sus créditos.

Un análisis basado en árboles ofrece algunas características atractivas, permite identificar grupos homogéneos, facilita la construcción de reglas para realizar pronósticos sobre casos individuales.

### **3. Desarrollo**

En esta parte se realiza todo el proceso de descubrimiento de conocimiento - KDD, siguiendo los pasos definidos en Ilustración 1, Primero empezamos con la presentación de las herramientas de análisis usadas para realizar el trabajo técnico y de análisis. Segundo se analiza la estructura de los datos de las encuestas de Hogares, estudiando la ficha técnica, Diccionario de datos y Cuestionarios disponibles en la descarga conjunta con la base de datos, Tercero se realiza el proceso de extracción y Limpieza de datos aquí extraemos los datos y lo importamos en una base de datos, para luego realizar la limpieza e integración y cargarlo en una tabla analítica. En el Cuarto paso, se procede realizar el análisis respectivo, que consiste en dos fases, la primera es seleccionar las 5 variables más influyentes del top 4 instituciones con mayor proporción de “Nada de confianza”, y la segunda parte es analizar el comportamiento de la relación de las variables influyentes con la confianza de cada institución seleccionada.

#### **3.1. Herramientas de Análisis**

##### **SQL SERVER 2012**

Es un Sistema Gestor de Base de datos de propiedad de Microsoft, la versión express es de uso libre y se ajusta a este trabajo, las bases que se descargan están en diferentes archivos, este software servirá para la importación y el procesamiento de los datos, el mismo que nos permite usar lenguaje SQL para realizar la transformación e integraciones de las base de datos de los estudios seleccionados, todo este proceso se llama (Extracción, Transformación y Carga) ETL parte del proceso de descubrimiento del conocimiento en minería de datos.

##### **IBM SPSS STATISTICS**

Este software es para realizar análisis estadístico y de minería de datos, será usado para obtener resultados de distribuciones y ejecución de algoritmos de minería de datos como Arboles de decisión y Análisis Factorial de Correspondencias, como también la Representación de Figuras usados en este trabajo.



### 3.2. Análisis de la estructura de los datos

Los datos usados son de bases de datos extraídas del repositorio de libre descarga <http://inei.inei.gob.pe/microdatos/> del Instituto Nacional de Estadística e Informática(INEI), cuenta con datos sobre las investigaciones del 2004 al 2014, cada base de datos está en formato .sav y .dbf, también se cuenta con la ficha técnica del estudio, diccionario de la base de datos y Cuestionarios usados para cada año y módulo respectivo como se ve en la tabla 2, cada módulo cuenta una cantidad de variables, en este trabajo solo se usará parte de las opiniones sobre confianza en las instituciones del gobierno correspondientes al Modulo con Código N° 85 llamado Gobernabilidad, Democracia y Transparencia.

De las bases de datos se extraerán solo las variables más relevantes por cada concepto de estudio.

En la tabla 3 se ven las cantidades de entrevistados, según Viviendas y Hogares desde el 2007 al 2014, este es un resultado obtenido luego de realizar el procesamiento con la base de datos descargada.

**Tabla 2:** Módulos seleccionados sobre la encuesta nacional de hogares ENAHO 2014

Nro	Año	Período	Código Encuesta	Código Módulo	Módulo
1	2014	55	440	1	<a href="#">Características de la Vivienda y del Hogar</a>
2	2014	55	440	2	<a href="#">Características de los Miembros del Hogar</a>
3	2014	55	440	3	<a href="#">Educación</a>
4	2014	55	440	4	<a href="#">Salud</a>
5	2014	55	440	5	<a href="#">Empleo e Ingresos</a>
6	2014	55	440	8	<a href="#">Instituciones Benéficas</a>
7	2014	55	440	10	<a href="#">Transportes y Comunicaciones</a>
8	2014	55	440	11	<a href="#">Servicios a la Vivienda</a>
9	2014	55	440	12	<a href="#">Esparcimiento , Diversión y Servicios de Cultura</a>
10	2014	55	440	15	<a href="#">Gastos de Transferencias</a>
11	2014	55	440	37	<a href="#">Programas Sociales (Miembros del Hogar)</a>
12	2014	55	440	77	<a href="#">Ingresos del Trabajador Independiente</a>
13	2014	55	440	78	<a href="#">Bienes y Servicios de Cuidados Personales</a>
14	2014	55	440	85	<a href="#">Gobernabilidad, Democracia y Transparencia</a>

Fuente. INEI (Instituto Nacional de estadísticas e informática)

**Tabla 3:** Distribución de la Cantidad de Viviendas y Hogares Encuestadas del 2007 al 2014, procesado según los datos disponibles en el repositorio del INEI

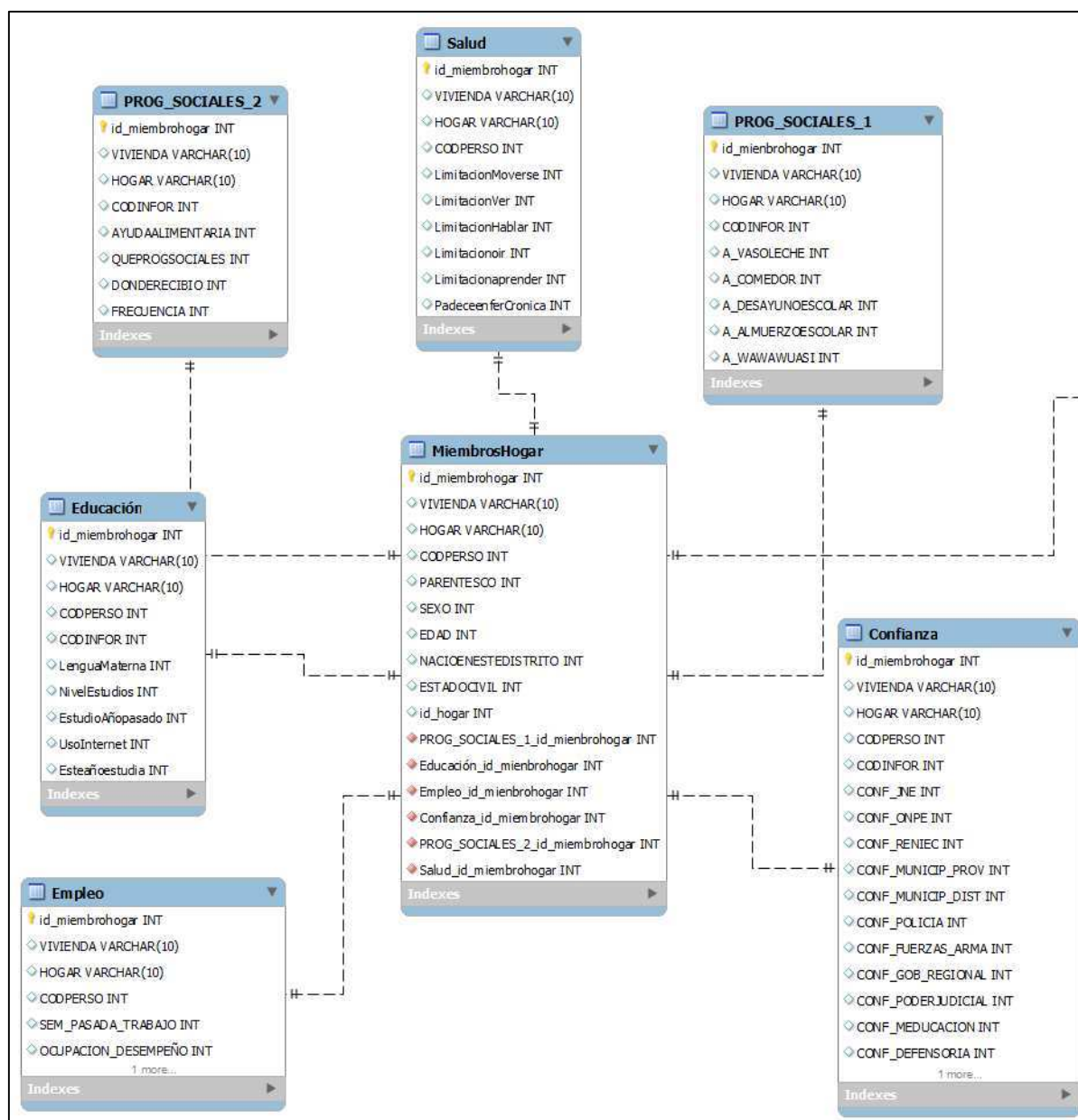
	2007	2008	2009	2010	2011	2012	2013	2014
Viviendas	25.947	25.425	26.058	26.684	31.956	32.051	39.109	39.646
Hogares	26.527	26.010	26.598	27.176	32.519	32.546	39.676	40.125

Elaboración propia, con datos del INEI

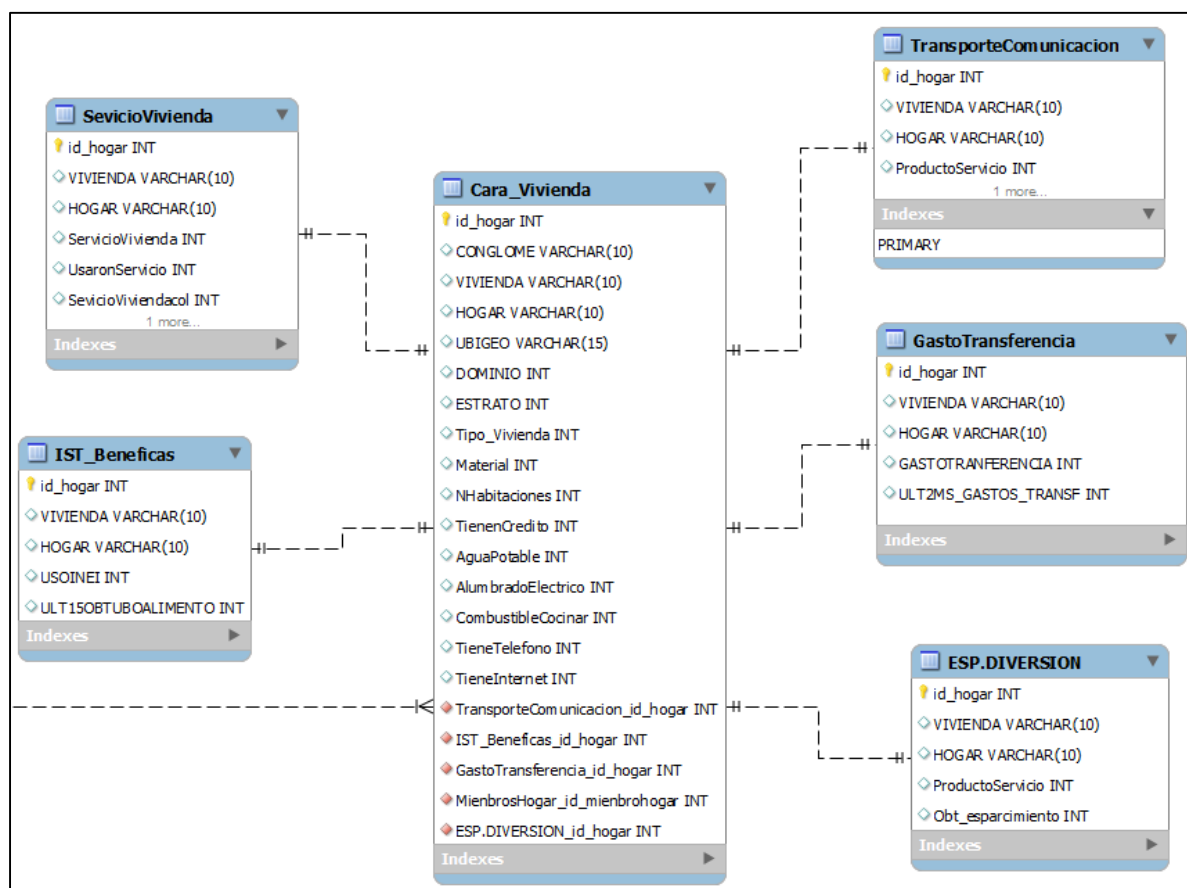
### 3.2.1. Diseño de la base de datos

Para importar los datos, se diseñó una base de datos en Sql Server, que se ve en la ilustración 4 y 5 mediante el diagrama relacional, el primer DER representa la relación de las características del miembro del hogar, como son educación, programas sociales, Salud, Empleo y el objeto de estudio, confianza. En el estudio de percepción de confianza solo son entrevistadas las personas mayores de edad. El segundo DER está relacionado a la vivienda o características que afectan a todos los miembros que viven en la vivienda, servicios que cuenta la vivienda, instituciones de beneficencia, servicio de transporte, diversión.

En el modelo DER, la mayoría de variables tienen el tipo de INT, debido que se almacenan los datos codificados, por ejemplo, Sexo (Masculino=1, Femenino=2), se puede revisar el diccionario de la base de datos, donde se especifica la correspondencia del código con su etiqueta y tenerlo en consideración. En todo el proceso de integración y limpieza se trabaja con datos codificados hasta obtener los datos integrados en la tabla analítica, luego estos datos se exportan a la base de SPSS que tienen las etiquetas respectivas de las variables.



**Ilustración 4:** DER de Características de los miembros del hogar.

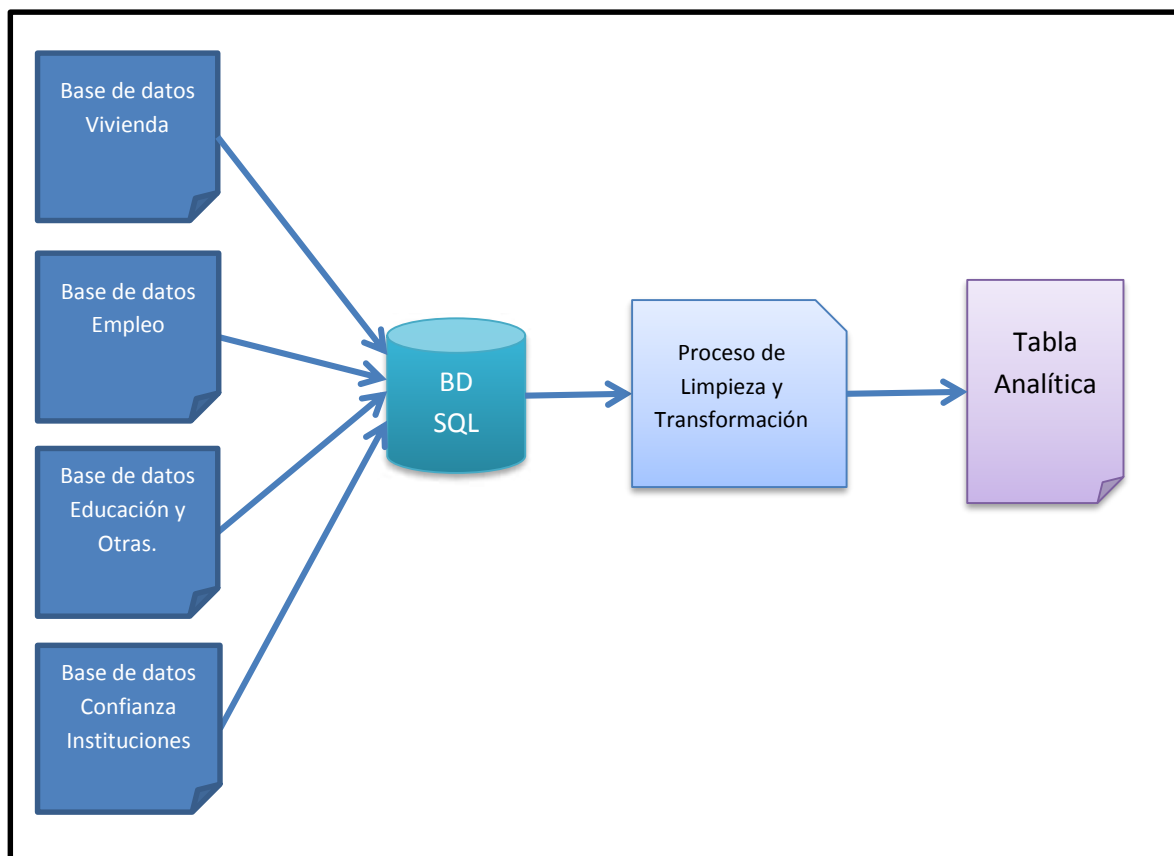


**Ilustración 5:** DER de Características de la Vivienda.

### 3.3. Etapa de Extracción Transformación y Carga(ETL)

En este proceso se integran los pasos de 1 al 4 del proceso KDD, aplicando a las bases de datos de los estudios seleccionados y luego se seleccionan las variables en un solo tablón analítico. Son 12 bases de datos diferentes que se descargan del repositorio, En la tabla 2 se muestra la lista de módulos, cada módulo trae los datos en archivos con extensión .sav junto con la documentación referidas al diccionario de la base de datos y la Ficha técnica, donde también se indican el proceso del estudio de investigación, como también trae los cuestionarios usados para las entrevistas.

Estas bases son exportados a una base de datos en SQL Server, donde se procede a realizar el proceso de ETL(Extracción Transformación y Carga) mediante script SQL, que permite consolidar en un solo archivo llamada tabla Analítica, en la Ilustración 6 se ve el modelo de ETL usado.

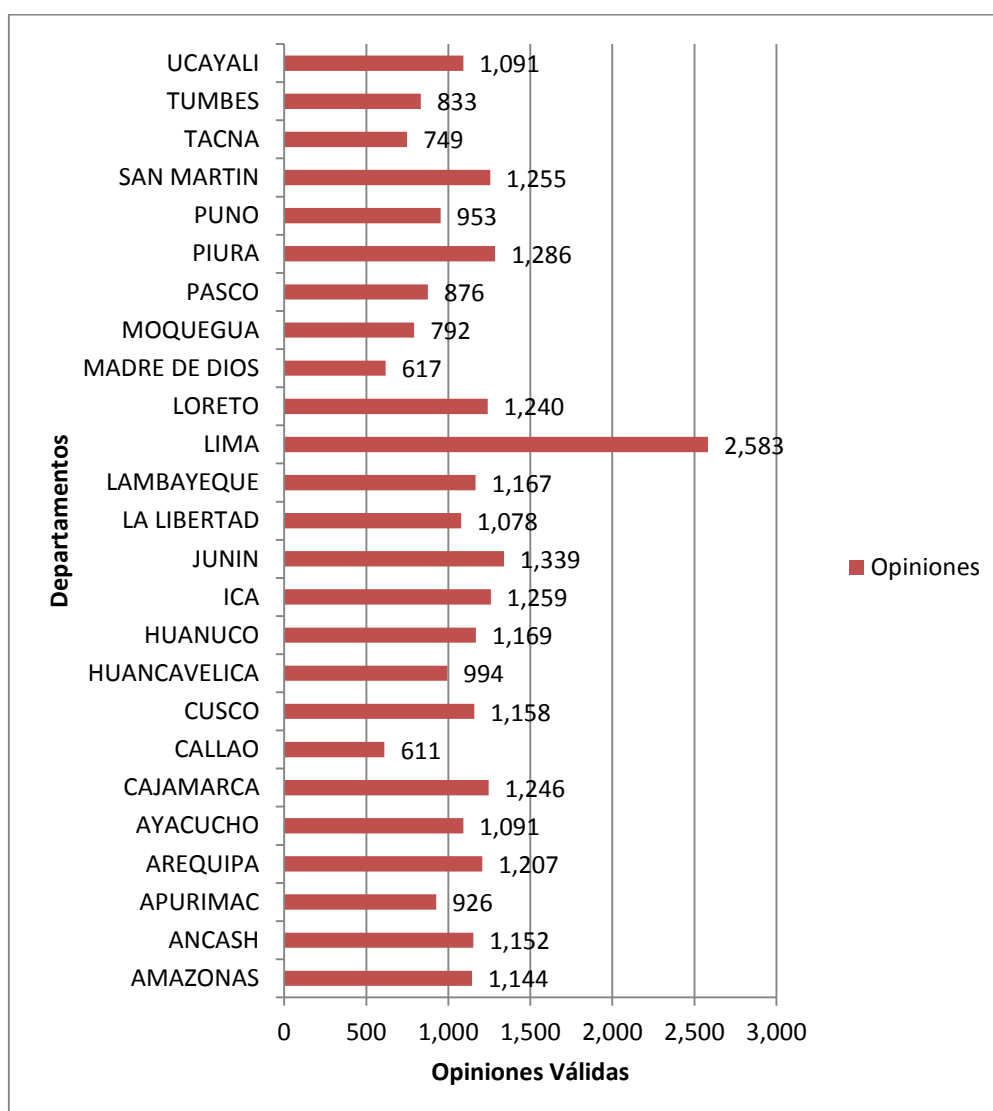


**Ilustración 6:** Proceso de ETL para la creación de la tabla analítica

### 3.4. Proceso de Exploración de datos y selección de variables

En esta parte se realiza el análisis exploratorio de todas las variables disponibles, para cada variable disponible, se realiza un análisis descriptivo y verifica el porcentaje de missing, si este cuenta con más del 25% se elimina del análisis, también se eliminan del análisis variables categóricas con mucha variabilidad, también variables que no proporcionan valor en algún análisis, por ejemplo variables del hogar: [se encuentra ausente del hogar 30 días o más], [la semana pasada estuvo trabajando] y solo se considera la variable trabaja (Si/No), en el módulo **educación** ( [centro de estudios] y [carrera universitaria que ha estudiado] son variables con mucha dispersión para el análisis y se excluye, [cual es el parentesco con el jefe del hogar] no aportan y se excluye), módulo salud (cuales son las razones por las que no tiene dni), empleo ( N° de personas que trabajan en su negocio, cuantas horas trabajo la semana pasada, ud laboró en su ocupación secundaria) se excluyen, en casos de variables de programas sociales, solo se consideró las variables que indican si es beneficiario o no, de esta forma se seleccionan las variables que serán usadas para el análisis de minería de datos.

El análisis inicia realizando las distribuciones generales sobre la percepción de la confianza del año 2014.



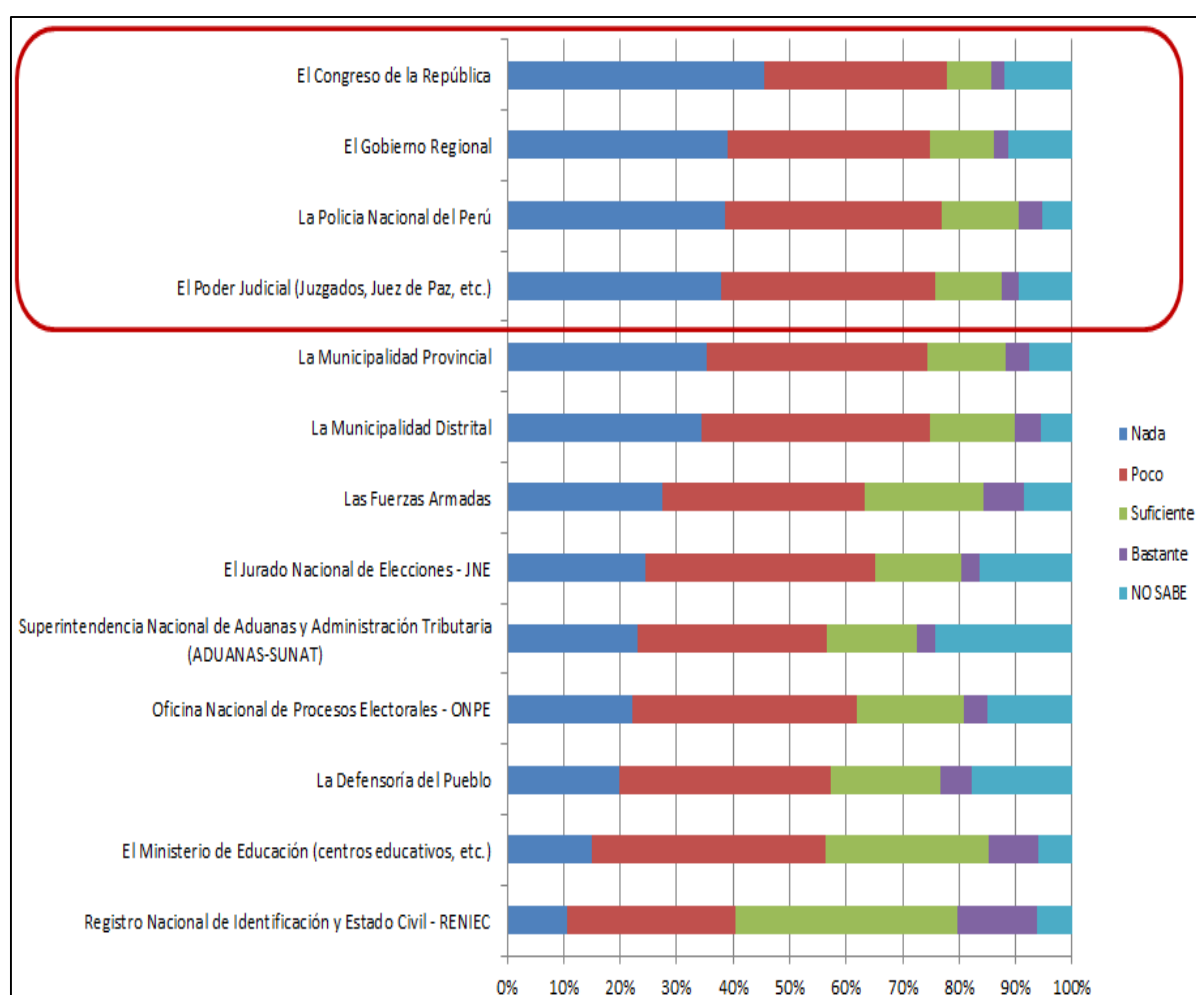
**Ilustración 7:** Distribución de opiniones validas según Departamentos del año 2014

Usamos los datos de la base del 2014, lo interesante para el gobierno es tratarlos por segmentos geográficos de la población, lo más manejable para este estudio es verlo por departamentos, el Perú está formado por 24 departamentos y la Provincia Constitucional del Callao, la ilustración 7 se ve la distribución de los datos encontrados, Lima la capital tiene la mayor cantidad de entrevistados con 2.583 y la menor cantidad es el Callao con 611.

**Tabla 4:** Distribución de Confianza con las instituciones del Gobierno 2014

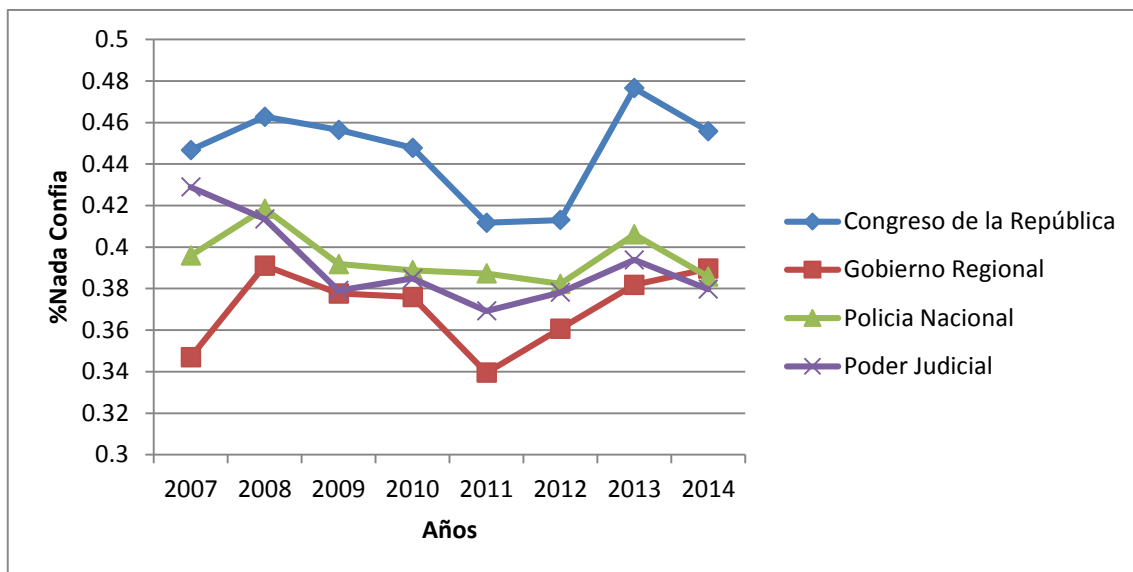
N°	Variable	Nombre de la Institución	Valoración de la Confianza							
			Nada	Poco	Suficiente	Bastante	NO SABE	Total Validos	Perdidos	Total
1	P1\$12	El Congreso de la República	12.679	8.991	2.189	589	3.368	27.816	3.032	30.848
2	P1\$08	El Gobierno Regional	10.838	10.007	3.092	748	3.131	27.816	3.032	30.848
3	P1\$06	La Policía Nacional del Perú	10.730	10.638	3.824	1.166	1.458	27.816	3.032	30.848
4	P1\$09	El Poder Judicial (Juzgados, Juez de Paz, etc.)	10.562	10.481	3.277	851	2.645	27.816	3.032	30.848
5	P1\$04	La Municipalidad Provincial	9.851	10.850	3.862	1.124	2.129	27.816	3.032	30.848
6	P1\$05	La Municipalidad Distrital	9.555	11.274	4.195	1.247	1.545	27.816	3.032	30.848
7	P1\$07	Las Fuerzas Armadas	7.656	9.936	5.839	1.999	2.386	27.816	3.032	30.848
8	P1\$01	El Jurado Nacional de Elecciones - JNE	6.773	11.319	4.243	929	4.552	27.816	3.032	30.848
9	P1\$20	Superintendencia Nacional de Aduanas y Administración Tributaria (ADUANAS-SUNAT)	6.424	9.293	4.463	867	6.769	27.816	3.032	30.848
10	P1\$02	Oficina Nacional de Procesos Electorales - ONPE	6.147	11.060	5.285	1.178	4.146	27.816	3.032	30.848
11	P1\$11	La Defensoría del Pueblo	5.538	10.371	5.444	1.526	4.937	27.816	3.032	30.848
12	P1\$10	El Ministerio de Educación (centros educativos, etc.)	4.179	11.500	8.035	2.452	1.650	27.816	3.032	30.848
13	P1\$03	Registro Nacional de Identificación y Estado Civil - RENIEC	2.961	8.255	10.920	3.966	1.714	27.816	3.032	30.848

En la base de datos del estudio sobre Gobernabilidad y transparencia se cuenta con 13 variables referidas a la confianza de instituciones que permanecen desde el año 2007 al 2014, en la tabla 4 se ve la distribución de las diferentes niveles de percepciones de cada institución, ordenados según la mayor cantidad de casos en percepción “Nada de Confianza”, la ilustración 8 también se ve su comportamiento proporcional, El congreso de la república está en el primer lugar con una alta percepción de “Nada de confianza”, Segundo está el Gobierno Regional, Tercero la Policía Nacional, así sucesivamente hasta el último que es Registro Nacional de identificación y Estado Civil-RENIEC, para este trabajo se eligen el top 4 de instituciones.



**Ilustración 8:** Distribución de Confianza con las instituciones del Gobierno 2014





**Ilustración 9:** Proporción de “Nada confía” del Top 4 de instituciones del 2007 al 2014

En la ilustración 9 se ve el comportamiento de “Nada de confianza” en las 4 instituciones seleccionadas del 2007 al 2014, se puede apreciar que el congreso de la república esta con la mayor proporción de “Nada de confianza”, del 2007 hasta el 2014, Luego se ve que la policía nacional esta segundo en los años 2008 al 2013 y el Gobierno regional toma su posición en el 2014, pero permanece cuarto desde el 2007 al 2013. El poder judicial se mantiene tercero durante los años 2008 al 2013.

**Tabla 5:** Proporciones de la Percepción de la confianza del Congreso de la República, por departamentos

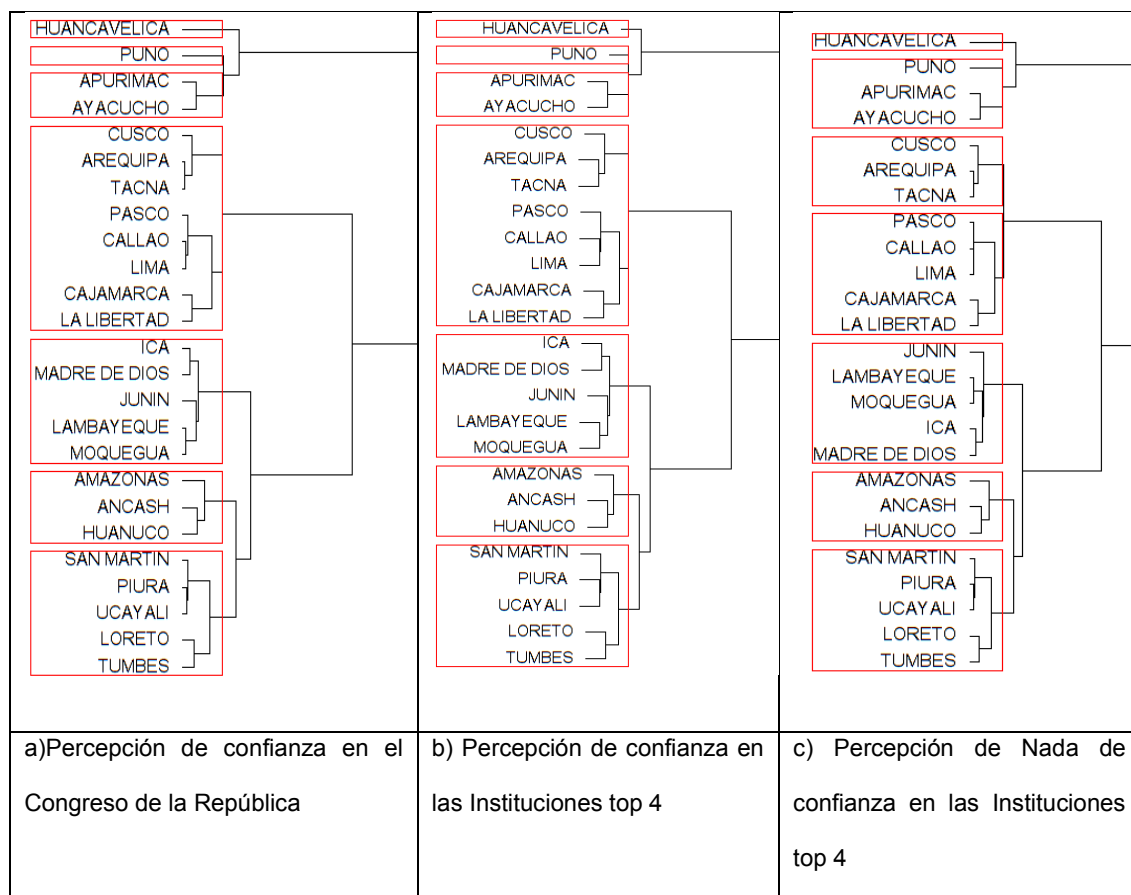
Percepción de la confianza en el Congreso de la República						
Departamentos	Nada	Poco	Suficiente	Bastante	NO SABE	TOTAL
AMAZONAS	33,30%	31,47%	10,49%	3,23%	21,50%	100%
ANCASH	39,58%	36,02%	7,12%	1,91%	15,36%	100%
APURIMAC	35,96%	27,11%	2,38%	3,02%	31,53%	100%
AREQUIPA	61,89%	23,86%	6,63%	0,99%	6,63%	100%
AYACUCHO	30,52%	25,48%	3,12%	2,66%	38,22%	100%
CAJAMARCA	50,24%	26,65%	11,96%	1,44%	9,71%	100%
CALLAO	58,27%	30,77%	8,02%	0,16%	2,78%	100%
CUSCO	60,54%	25,04%	3,63%	1,64%	9,15%	100%
HUANCAVELICA	18,41%	20,82%	6,64%	0,60%	53,52%	100%
HUANUCO	46,11%	30,11%	4,70%	3,17%	15,91%	100%
ICA	51,07%	39,79%	7,39%	0,64%	1,11%	100%
JUNIN	44,29%	40,85%	6,27%	0,82%	7,77%	100%
LA LIBERTAD	49,44%	31,82%	9,74%	1,48%	7,51%	100%
LAMBAYEQUE	47,30%	37,19%	10,03%	1,11%	4,37%	100%
LIMA	57,14%	29,89%	7,59%	0,81%	4,57%	100%
LORETO	33,47%	43,79%	11,94%	7,26%	3,55%	100%
MADRE DE DIOS	49,27%	38,25%	6,00%	3,89%	2,59%	100%
MOQUEGUA	46,09%	38,01%	11,74%	0,76%	3,41%	100%
PASCO	56,62%	31,05%	5,59%	2,28%	4,45%	100%
PIURA	38,26%	37,87%	11,82%	2,95%	9,10%	100%
PUNO	50,58%	14,48%	1,36%	0,94%	32,63%	100%
SAN MARTIN	38,41%	37,37%	9,80%	5,10%	9,32%	100%
TACNA	62,88%	23,50%	7,21%	1,07%	5,34%	100%
TUMBES	38,30%	45,14%	12,73%	2,28%	1,56%	100%
UCAYALI	36,76%	39,87%	11,00%	3,02%	9,35%	100%

La tabla 5 se ve las proporciones de percepción, por departamentos, esto para analizar en los diferentes departamentos, usamos las proporciones de opiniones referidas a los diferentes niveles, Nada, Poco, Suficiente, Bastante y No Sabe, estos valores están en rangos de 0 –100.

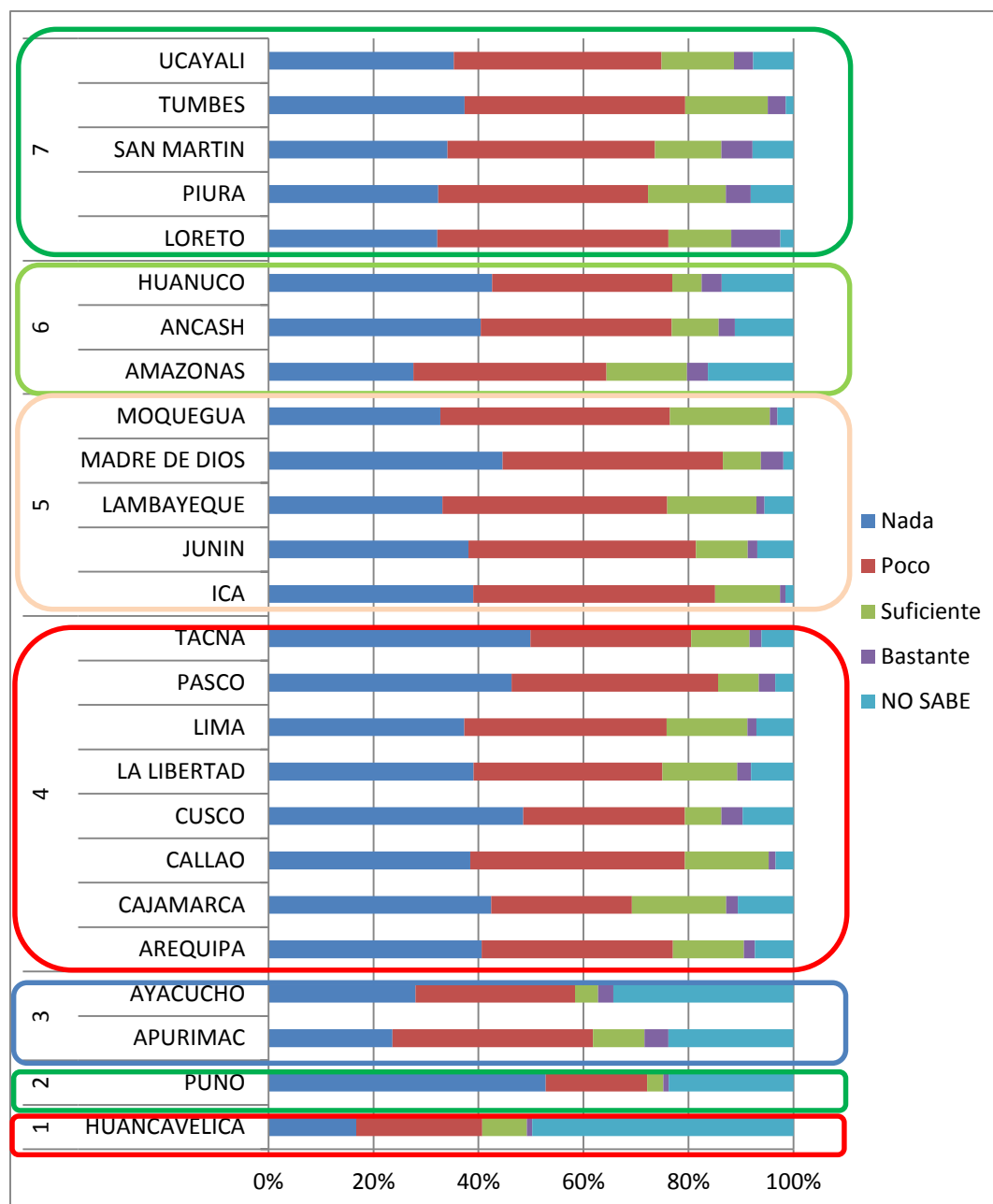
La Ilustración 10, se ve el comportamiento de afinidades de opiniones por departamentos, las afinidades fueron determinados realizando los siguientes criterios: **“a) Percepción de confianza en el Congreso de la República”**, para este caso se usaron todos los niveles de proporciones referidos al Congreso de la República en 5 vectores. **“b) Percepción de confianza en las instituciones Top 4”**, para este caso se usaron todos los niveles de confianza de las 4 instituciones, obteniendo 20 vectores. **”C) Percepción de**

**Nada de confianza en las Instituciones top 4**", para este caso solo se usó los niveles de **"Nada confía"** de las 4 Instituciones, obteniendo 4 variables. En cada caso se usaron los vectores para ejecutar el algoritmo de agrupamiento jerárquico, de esta forma determinar los grupos de afinidades.

Los resultados muestran que los grupos de departamentos tienen una afinidad en las proporciones de opiniones, en particular las jerarquías de a y b coinciden en un 100%, el grupo c muestra diferencias como es que el departamento de PUNO se integra al grupo de Apurímac y Ayacucho, mientras que el grupo donde pertenece CUSCO grupo a y b se divide en dos grupos en c, el departamento de Huancavelica no tiene afinidad a ningún departamento en todos los casos.

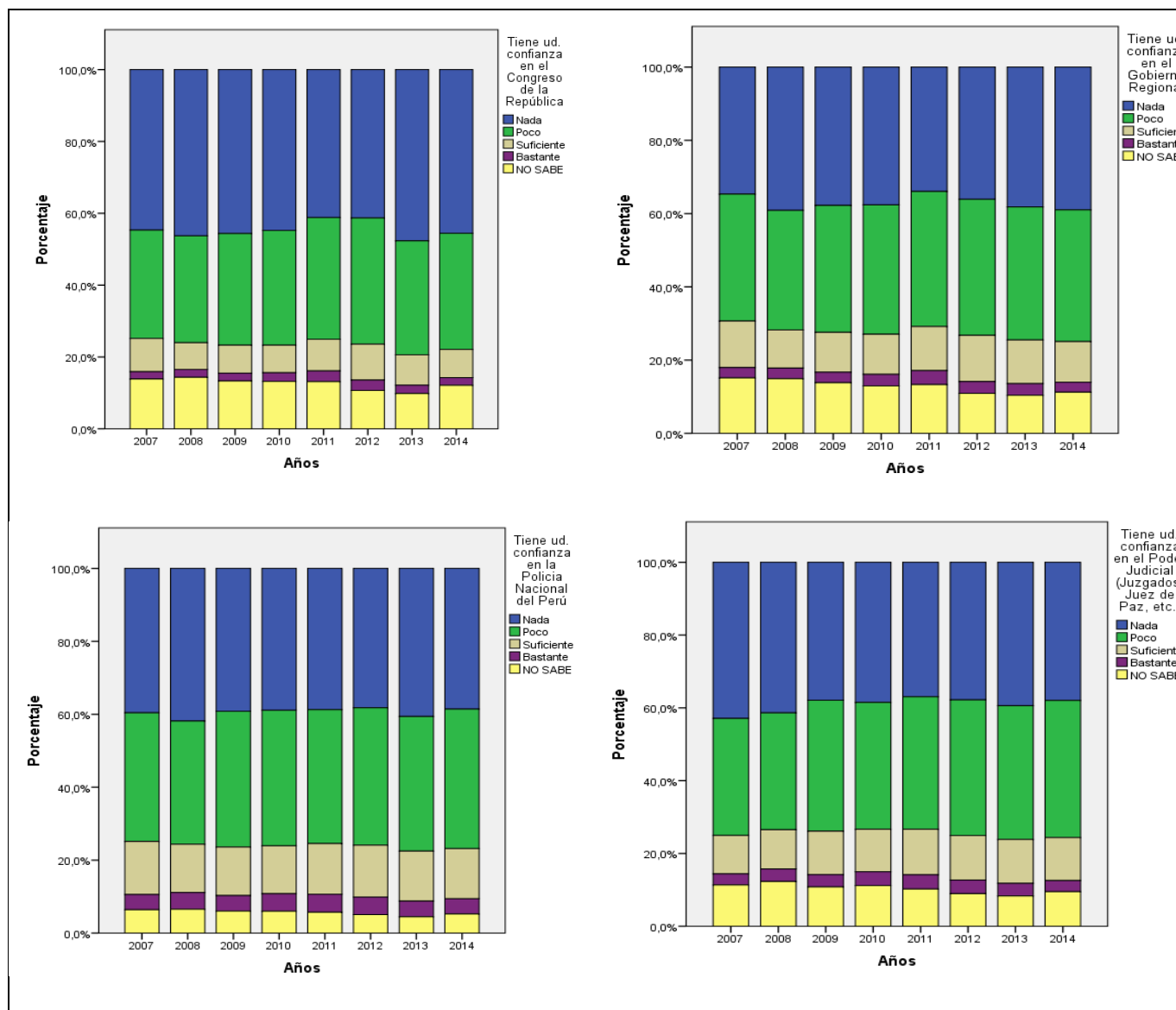


**Ilustración 10:** Afinidad por Departamentos según la percepción de confianza del 2014



**Ilustración 11:** Percepción del top 4 Instituciones en segmentos similares por departamentos

La ilustración 11 se desarrolló usando los resultados de promedios ponderados de los vectores del mismo nivel de percepción de las 4 Instituciones, con las mismas variables usadas en la Ilustración 10 – caso “**b) Percepción de confianza en las instituciones Top 4**”. Los resultados muestran a Puno y Huancavelica como dos departamentos que tienden a ser no afines a los demás departamentos, que si pertenecen a un grupo de afinidad, por ejemplo Puno es el departamento con mayor proporción de “Nada de confianza”, más del 50%, el otro departamento es Huancavelica, tiene mayor proporción en “No sabe” sobre la institución.



**Ilustración 12:** Distribución de Top 4 Instituciones, del 2007 al 2014

En la ilustración 12 se ve el comportamiento de la percepción de la confianza de las 4 instituciones con mayor porcentaje de “No Confianza”, durante el 2007 al 2014, se ve que no hay variaciones drásticas en la percepción.

**Tabla 6:** Resumen de porcentajes del Top 4 de las instituciones Seleccionadas

Variable	Nombre de la Institución	Nada	Poco	Suficiente	Bastante	NO SABE	Total Validos
P1\$12	El Congreso de la República	45.6%	32.3%	7.9%	2.1%	12.1%	100%
P1\$08	El Gobierno Regional	39.0%	36.0%	11.1%	2.7%	11.3%	100%
P1\$06	La Policía Nacional del Perú	38.6%	38.2%	13.7%	4.2%	5.2%	100%
P1\$09	El Poder Judicial (Juzgados, Juez de Paz, etc.)	38.0%	37.7%	11.8%	3.1%	9.5%	100%
<b>Promedio %</b>		<b>40.3%</b>	<b>36.1%</b>	<b>11.1%</b>	<b>3.0%</b>	<b>9.5%</b>	<b>100.0%</b>

En la tabla 6, se detallan las proporciones en los diferentes niveles de Confianza en el top 4 de instituciones del gobierno, el congreso de la república es el primero con un 45,6% que “Nada Confía”, Esto se acerca al 50% el cual es un valor muy crítico para una institución que representa al pueblo peruano en los diferentes departamentos del Perú, el congreso de la república es el que lleva la voz del pueblo al parlamento para proponer leyes a su favor.

### 3.5. Selección de las variables más influyentes

En esta fase, se selecciona las variables más influyentes en la confianza, para cada institución seleccionada que se ve en la tabla 6, usando el algoritmo de árbol de decisión tipo Chaid, se obtienen las 5 variables más influyentes, la variable que está en la raíz es considerada la variable que aporta más información que las demás, por ende es la que influye más en la percepción de confianza. No se usará el árbol como modelo predictivo, sino para obtener las 5 variables más influyentes. Para ello se correrán 5 árboles de forma secuencial, el primer árbol con todas las variables, y como resultado obtenemos un árbol en donde obtenemos la primera variable influyente, Para obtener la Segunda variable corremos el segundo árbol con todas las variables menos la variable que salió como la más influyente en el primer árbol, de esta forma tenemos la segunda variable más influyente, tomando ese criterio se corren 5 árboles para cada una de las 4 instituciones seleccionadas.

Cada árbol obtenido, segmenta óptimamente agrupando las modalidades de cada variable, de esa forma aporta la mayor información en la variable confianza, se pondrán los árboles de modo explicativo.

En la tabla 7 se ven los resultados de las 5 variables más influyentes para cada institución, el *[nivel de estudios]* es la primera variable que influye en la percepción de confianza de las 4 instituciones públicas, como también se ve que *[lengua materna]* es la segunda variable más influyente, el *[dominio Geográfico]* es la tercera variable más influyente.

La cuarta variable más influyente cambia, entre el *[material de las paredes]* y *[material de piso]* que influye en la percepción de la confianza, poder judicial y gobierno regional comparten la variable de *[material de pisos]*, Congreso de la república con policía nacional tienen la variable de *[material predominante en las paredes exteriores]*.

Como quinta variable, que influye al congreso de la república tenemos al *[combustible que usan para cocinar en el hogar]*, las demás instituciones tienen la variable *[El hogar no tiene: teléfono fijo, celular, tv. Cable, Internet]*.

**Tabla 7:** Resumen de las variables más influyentes en la confianza sobre las 4 Instituciones

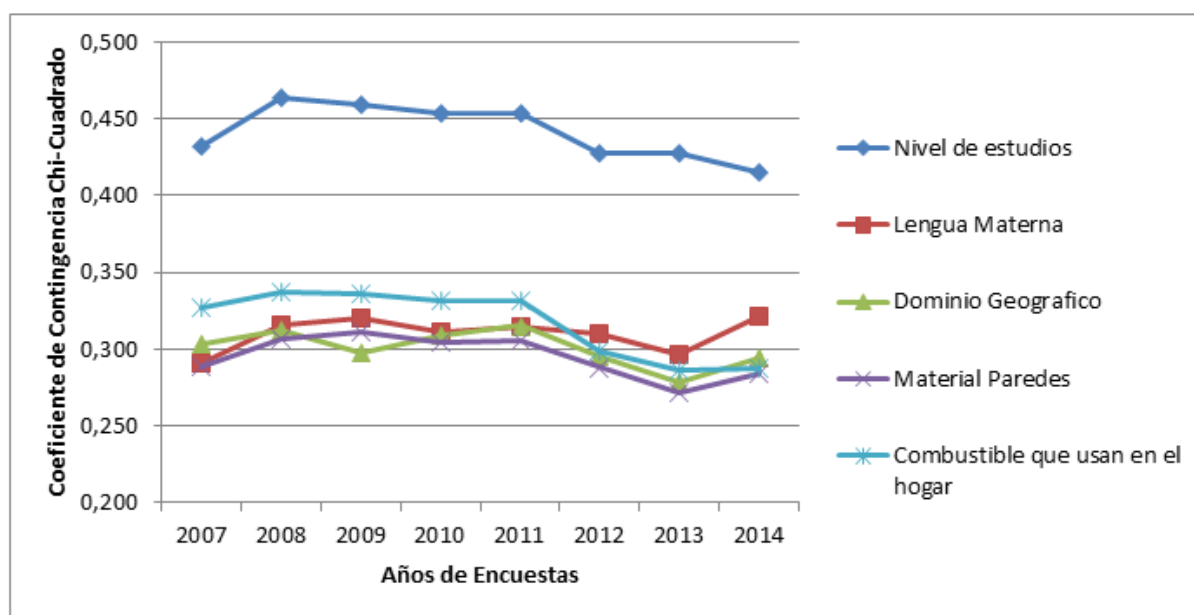
N° Árbol	Poder Judicial	Gobierno Regional	Congreso República	Policía Nacional
1	Nivel de estudios	Nivel de estudios	Nivel de estudios	Nivel de estudios
2	Lengua Materna	Lengua Materna	Lengua Materna	Lengua Materna
3	Dominio Geográfico	Dominio Geográfico	Dominio Geográfico	Dominio Geográfico
4	El material predominante en los pisos es:	El material predominante en los pisos es:	El material predominante en las paredes exteriores es	El material predominante en las paredes exteriores es
5	El hogar no tiene: teléfono fijo, celular, tv. cable, Internet	El hogar no tiene: teléfono fijo, celular, tv. Cable, Internet.	Combustible que usan en el hogar para cocinar: Mayor Frecuencia	El hogar no tiene: teléfono fijo, celular, tv. cable, Internet

### 3.5.1. Confianza en el Congreso de la República

En esta parte se detalla la evolución temporal de cómo influyen las 5 variables más influyentes en el 2014, en los años 2007 al 2014, se ve en la tabla 8, el resultado del coeficiente de contingencia durante el 2007 al 2014, como también gráficamente la ilustración 13.

**Tabla 8:** Resumen de las variables más influyentes en la confianza sobre el Congreso de la República

N°	Variables más Influyentes	Coeficiente de Contingencia de Chi cuadrado							
		2007	2008	2009	2010	2011	2012	2013	2014
1	Nivel de estudios	0,432	0,464	0,459	0,454	0,454	0,428	0,427	0,415
2	Lengua Materna	0,291	0,316	0,321	0,311	0,315	0,310	0,296	0,321
3	Dominio Geográfico	0,303	0,313	0,298	0,309	0,315	0,295	0,279	0,294
4	Material Paredes	0,289	0,306	0,311	0,304	0,306	0,288	0,272	0,284
5	Combustible que usan en el hogar	0,327	0,337	0,336	0,331	0,331	0,299	0,287	0,287



**Ilustración 13:** Evolución del coeficiente de contingencia sobre las variables más influyentes en la Confianza sobre el Congreso de la República, del 2007 al 2014.

Con los resultados obtenidos usando árbol chaid, tenemos 5 variables más influyentes en la percepción de confianza en el congreso de la república del 2014, Calculamos el coeficiente de contingencia del Chi cuadrado para ver el nivel de influencia de las variables y como es su comportamiento de influencia desde el 2007 al 2014. Los valores del indicador



están dentro del rango  $<0-0,5]$  considerado Asociación débil, sin embargo el análisis de la influencia a nivel valores de las variables prueba ser significativo.

La variable con mayor influencia es el [*nivel de estudios*], y se mantiene en esa posición desde el 2007 al 2014. le sigue [*Lengua materna*], que del 2012-2014 se mantiene como segundo más influyente, luego [*Material de paredes*] se mantiene del 2010-2014 como la quinta más influyente, las demás variables no permanecen estables.

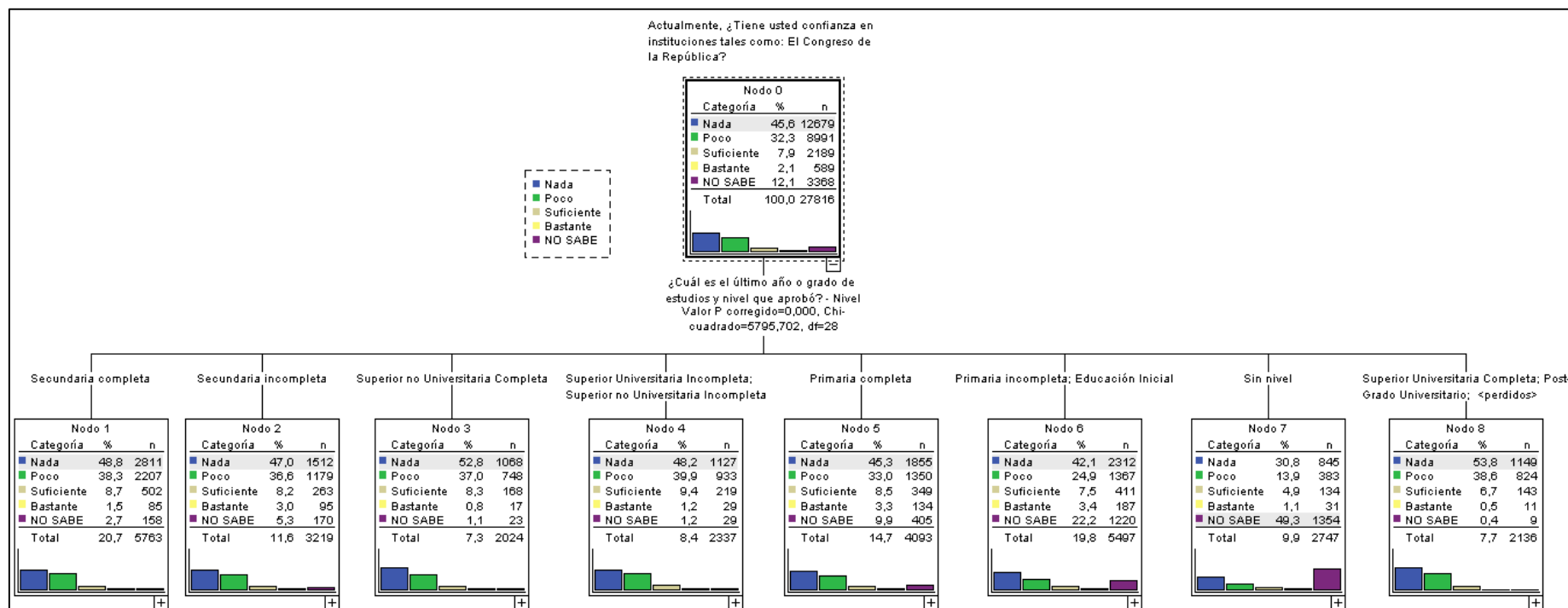
En la ilustración 14, se ve el primer árbol obtenido del algoritmo Chaid, usando todas las variables de entrada, se obtiene [*nivel de estudios*] como la variable más influyente y las segmentaciones de sus modalidades de esta variable que proporcionan mayor información en la influencia en la percepción sobre el congreso, las proporciones de opiniones sobre la percepción en sus diferentes grados, es fácil observar que el Nodo 7 contiene la mayor proporción de personas que no tienen algún estudio en su vida, y son necesariamente los que “No Saben” la existencia del Congreso de la República, dentro de nuestro objetivo, la percepción “Nada confía”, está muy distribuido homogéneamente en los diferentes niveles de estudios, aunque el nodo 8 da un mayor proporción 53.8% “Nada confía” son personas de “*Superior universitaria Completa*” y con “*Posgrado*”, se sospecha que hoy el congreso no es una institución donde el ciudadano bien preparado académicamente confíe plenamente.

En la ilustración 15, se ve el resultado del segundo árbol chaid, en su primer nivel, muestra la variable más influyente, este es el [*Lengua materna*], con las segmentaciones en las modalidades, son 4 grupos que aportan la máxima información en la variable de la percepción, lo más notable es que las lenguas Nativas como el, “*Aymara*”, “*Quechua*” y otras tienen alta proporción de “No Sabe”, esto se debe que hoy las comunicaciones en su mayor proporción llegan en el idioma castellano, también podemos observar que el grupo “*Quechua*” y “*Aymara*” tienen su distribución muy afín, pero difiere totalmente con los de habla “*Castellano*”.

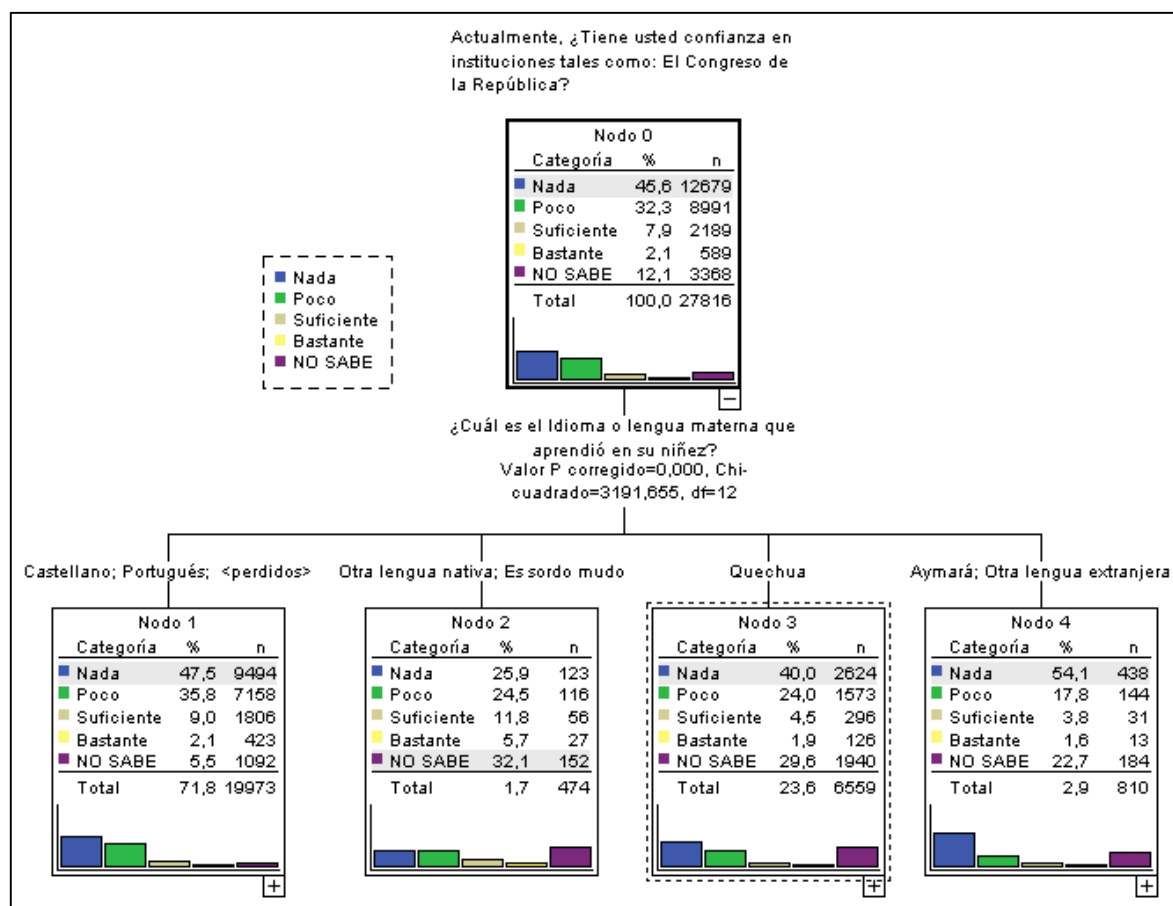
En la ilustración 16, se ve el resultado del tercer árbol chaid, sin las dos variables más influyentes obtenidas, esta vez obtenemos la tercera más influyente que es el [*Dominio geográfico*], siendo esta una variable de zonas de influencias entre las diferentes poblaciones, las zonas de la sierra tienen una distribución homogénea como se ve, y las de la costa se parecen mucho a lo de la selva. Se ven que en todas las zonas hay una alta proporción de “Nada de confianza,” pero si los de la Sierra tiene un porcentaje que difiere a los de la costa y selva en “No sabe sobre el congreso de la república”.

En la ilustración 17, se ve el resultado del Cuarto árbol chaid, sin las tres variables más influyentes obtenidas, esta vez obtenemos la cuarta variable más influyente, siendo [material de paredes], podría ser que esto da una forma de percepción de la población que vive en diferentes contextos de vivienda sienta diferente percepción en la confianza con el congreso de la república, se ve que la mayor representación está en “ladrillos y bloque de cemento” con alto porcentaje de 52,4% de “Nada de confianza”, se ve también que “No Sabe” de congreso de la república son segmentos de “Tapia, Piedra con Barro” y “adobe” y “piedra de sillar”.

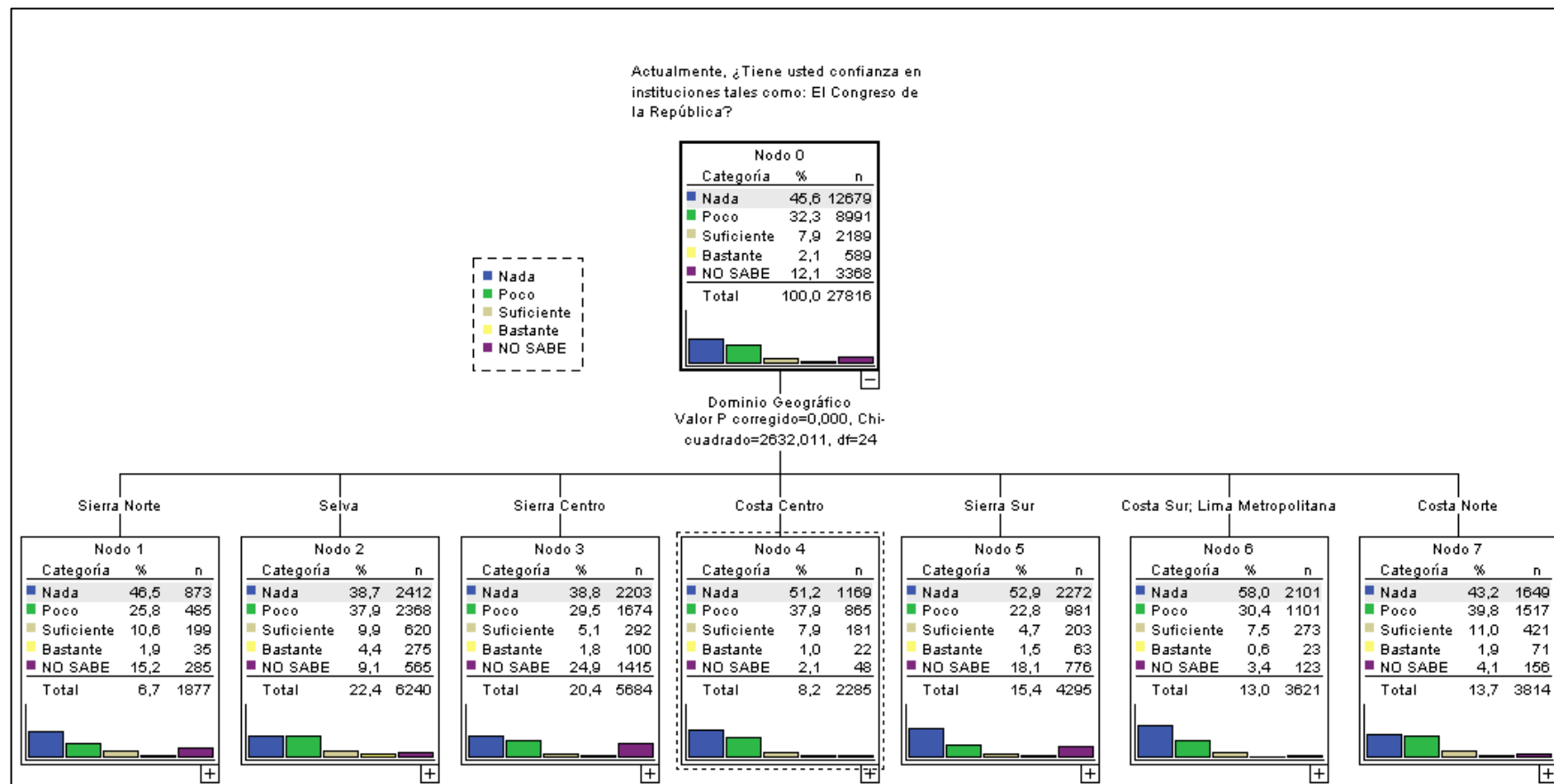
En la ilustración 18, se ve el resultado del Quinto árbol chaid, sin las 4 variables más influyentes obtenidas, esta vez obtenemos la quinta variable más influyente a [Combustible que usan en el hogar], si bien es cierto el más usado es el “GLP”(gas licuado de petróleo) y se ve que la gente que la consume tiene una percepción 50,6% de “Nada de confianza” y 35.8% Poco, una porción de la gente que “No Sabe” están en segmentos de uso de “Leña” y “otros combustibles”.



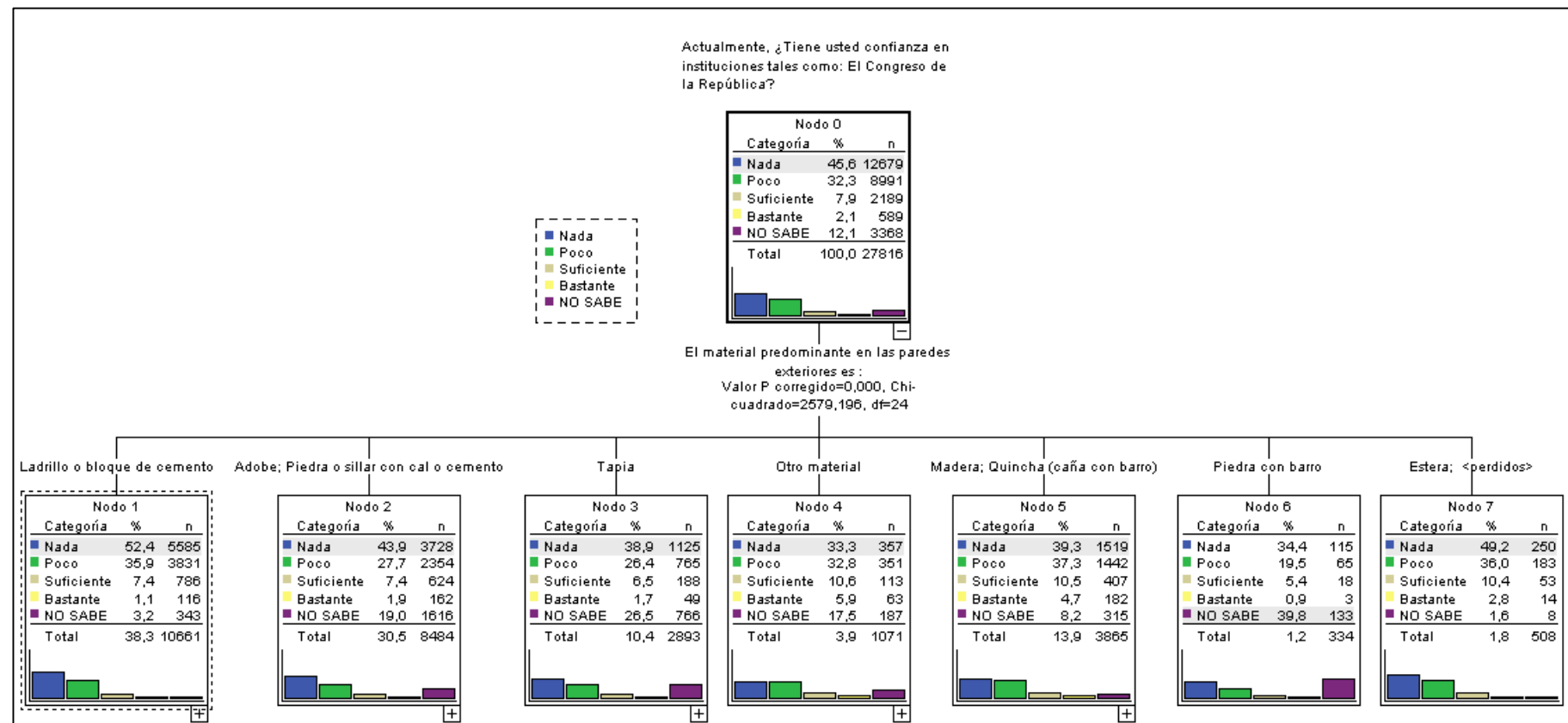
**Ilustración 14:** Primera variable más Influyente en la Confianza en el congreso de la República



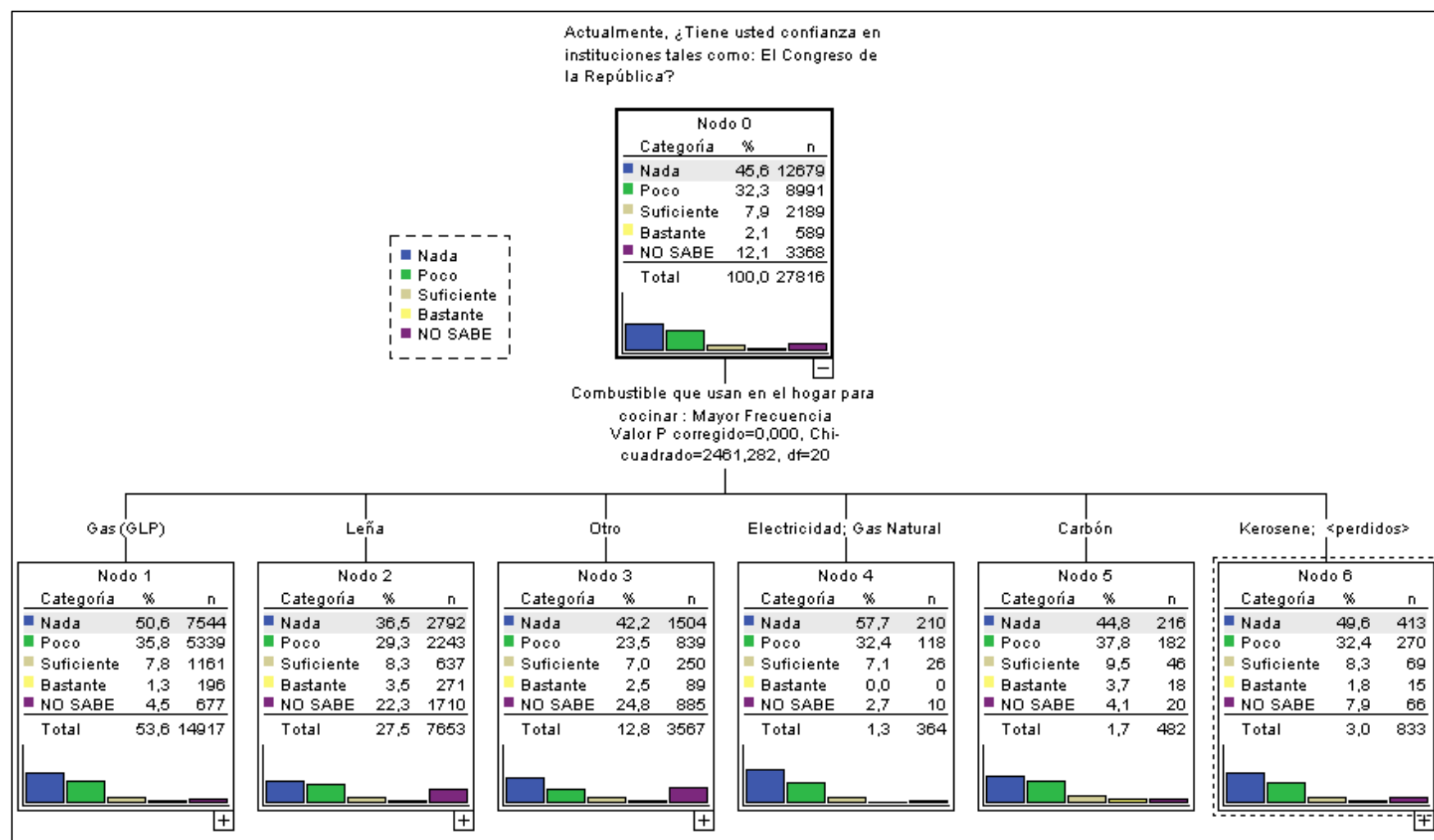
**Ilustración 15:** Segunda Variable más Influyente en la Confianza en el Congreso de la República



**Ilustración 16:** Tercera Variable más Influyente en la Confianza en el Congreso de la República



**Ilustración 17:** Cuarta Variable más Influyente en la Confianza en el Congreso de la República



**Ilustración 18:** Quinta Variable más Influyente en la Confianza en el Congreso de la República

En la ilustración 19, se ve en detalle la atracción que existe entre las modalidades del *[nivel de estudios]* y el *[nivel de confianza]* del congreso de la república, comparando el resultado de los años 2007, 2013 y 2014 en el 2007 se ve una atracción fuerte entre “Poco de confianza” con “secundario completo” y va bajando de atracción con niveles superiores de educación con mucha cercanía, y el 2014 tiene el mismo comportamiento, como también están atraídos por “Nada de Confianza”, en términos de solo esta modalidad se ve que mientras más nivel de estudios se cuente se va alejando la asociación.

Las personas sin estudios están más atraídas a No saben del congreso de la república, las que cuentan con un nivel bajo de estudios como nivel inicial y primaria incompleta están atraídas con bastante confianza, mientras más nivel de estudios alcance la confianza va bajando de nivel.

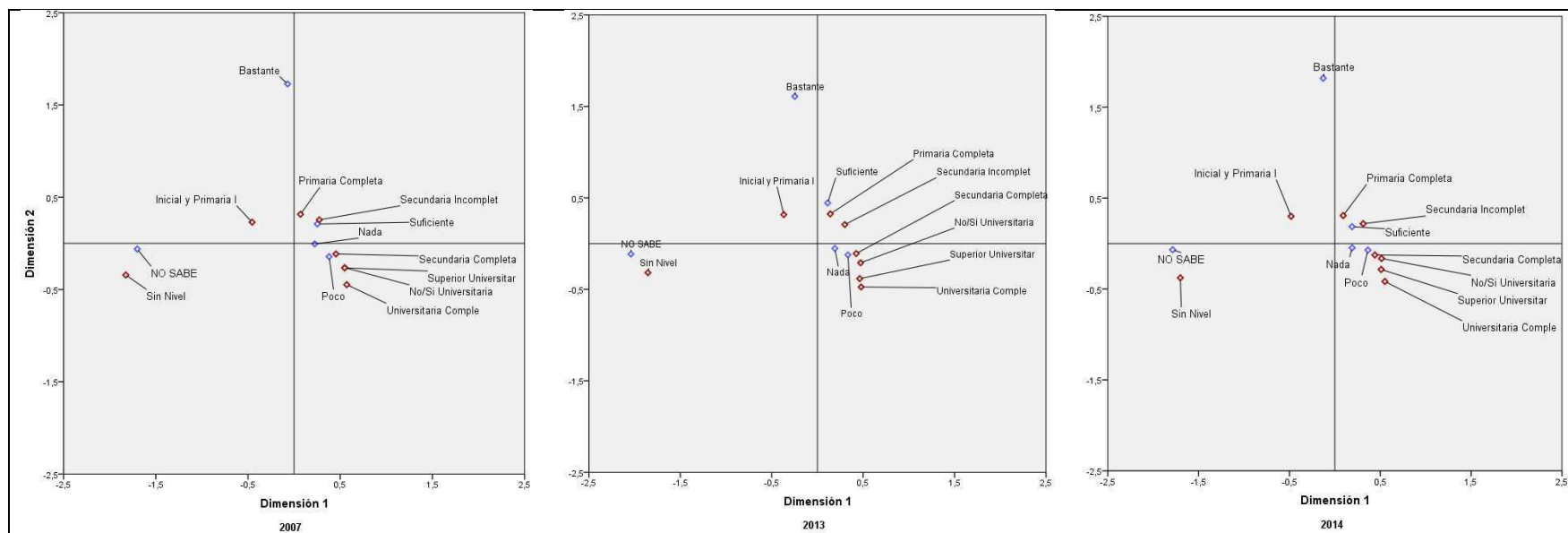
La ilustración 20 se ve la atracción entre las modalidades de la Lengua materna y Confianza del congreso de la república, una comparativa de 2014 con 2007 y 2013, la relación del 2014 se ve que hablantes de Quechua están más atraídos con “No Sabe” de Congreso de la república, los de Lengua “Aymara” se ve claro su relación, mientras que “Castellano” son más atraídos con “Poco Confía” y con menor atracción con “Nada” y “Suficiente” confianza.

La ilustración 21 se ve la atracción entre las modalidades de Dominio geográfico y confianza en el Congreso de la República, una comparación del 2014 con 2007 y 2013, los dominios de “Costa Norte” y “Selva” están más atraídos por “Poco” y “Suficiente” Confianza pero el 2007 “Selva” estaba menos atraída, luego “Costa centro” y “Costa sur/Lima metropolitana” más atraídas con “Nada” de confianza, Sierra Sur y Norte con una baja atracción por “Nada” de confianza, Sierra Centro más atraídos por No sabe de confianza y al pasar los años la relación se fortalece.

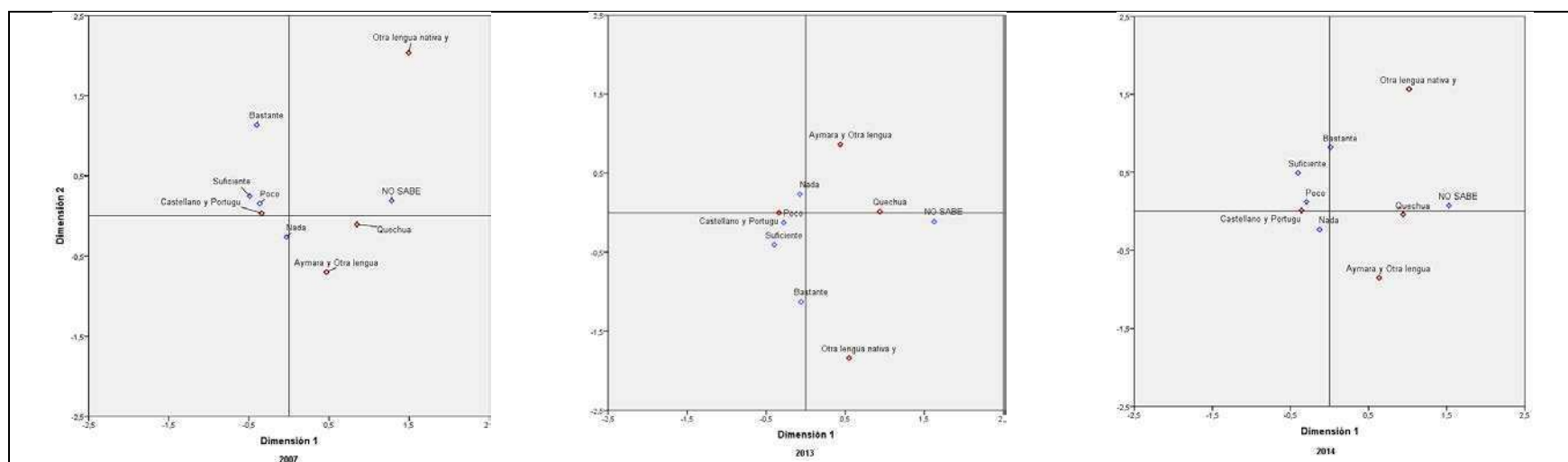
La ilustración 22, se ve la atracción entre las modalidades del Material predominante en las paredes exteriores de su vivienda y la percepción de la confianza en el congreso de la república, en sus modalidades un comparativo del 2014 con 2007 y 2013, 2007 y 2013 paredes de Piedra o Sillar están muy atraídos por Nada de confianza solo el 2014 la atracción se debilita, también es atraído por Ladrillo o bloque de cemento, Estera va cambiando su atracción en los años, “Quincha (caña con barro)” y “Madera” están más atraídos por “Suficiente confianza” desde el 2013, “Adobe / piedra con barro” y “Tapia” están más atraídos por “No Sabe”.



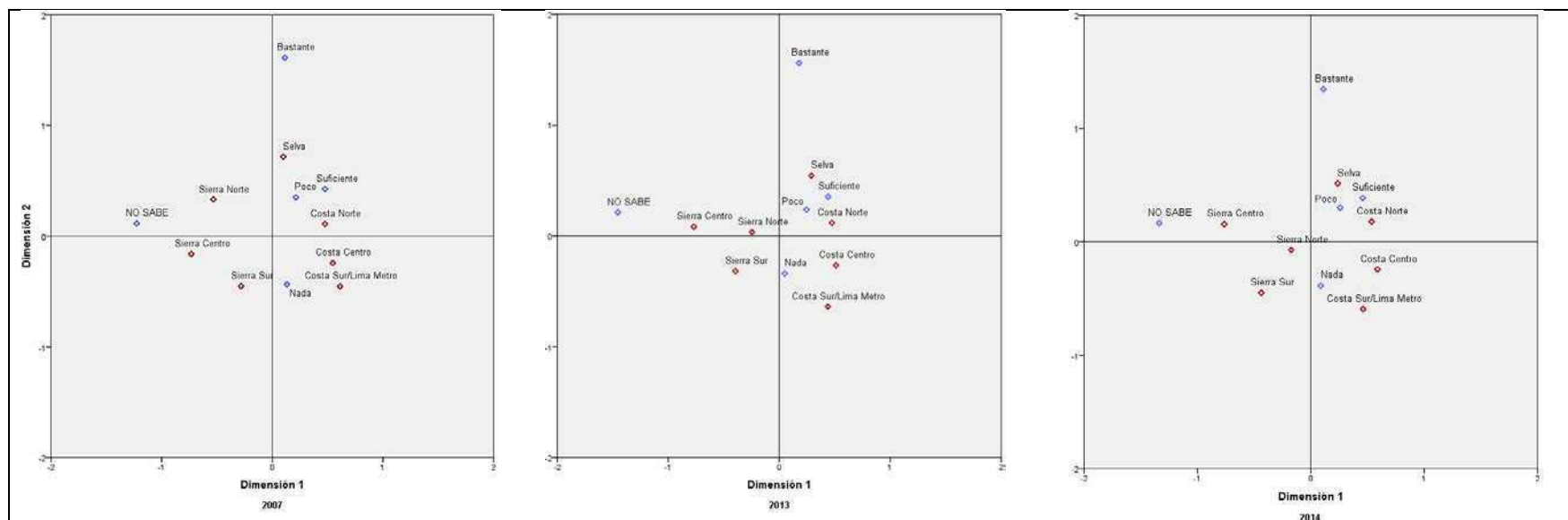
La ilustración 23 se ven las atracciones entre las modalidades del *[Combustible que usan en el hogar para cocinar]* con la confianza en el congreso de la república, aquí una comparativa del 2014 con 2007 y 2013. “Nada de confianza” está relacionado por el consumo de la “Electricidad y Gas Natural”, “Gas (GLP)” y “carbón” con “poco de confianza”, “No sabe” con otro tipo de combustible.



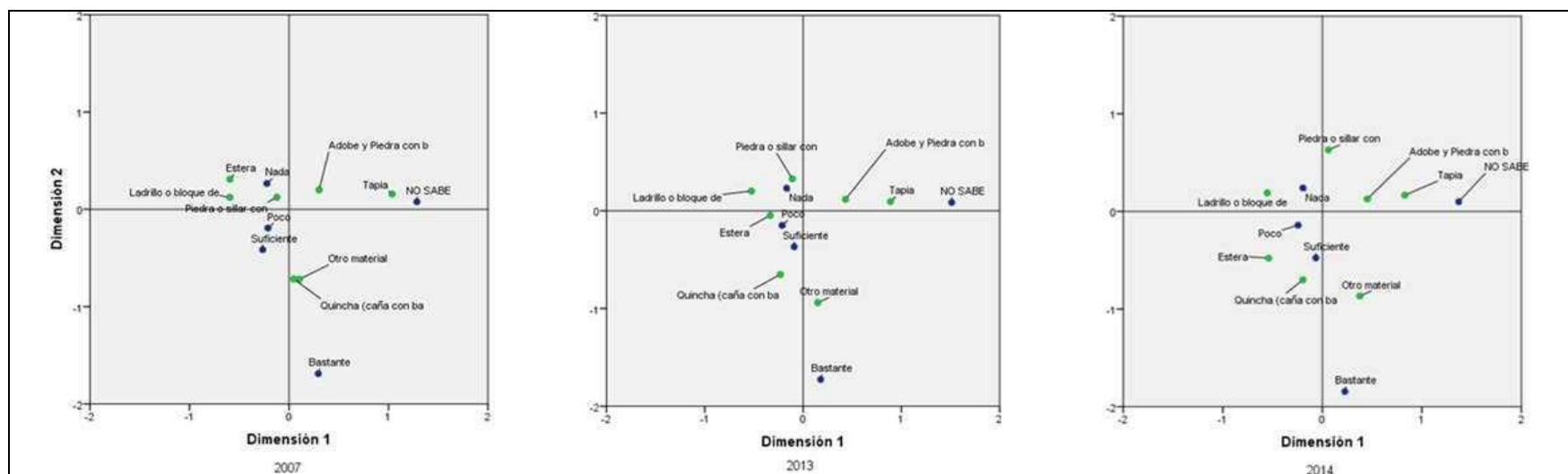
**Ilustración 19:** Nivel de estudios y la Confianza en el Congreso de la República, 2007, 2013 y 2014



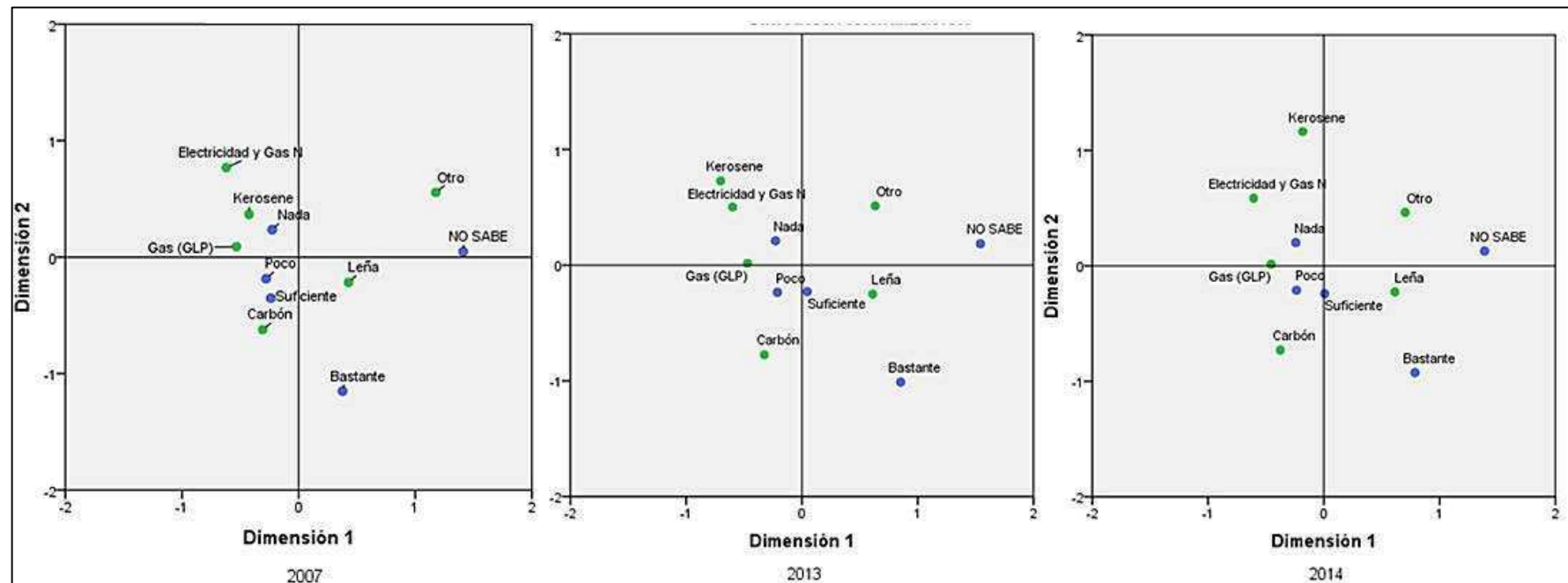
**Ilustración 20:** Lengua Materna y la Confianza en el Congreso de la República, 2007, 2013 y 2014



**Ilustración 21:** Dominio Geográfico y la Confianza en el Congreso de la República, 2007, 2013 y 2014



**Ilustración 22:** Material Predominante en las paredes y la Confianza en el Congreso de la República, 2007, 2013 y 2014



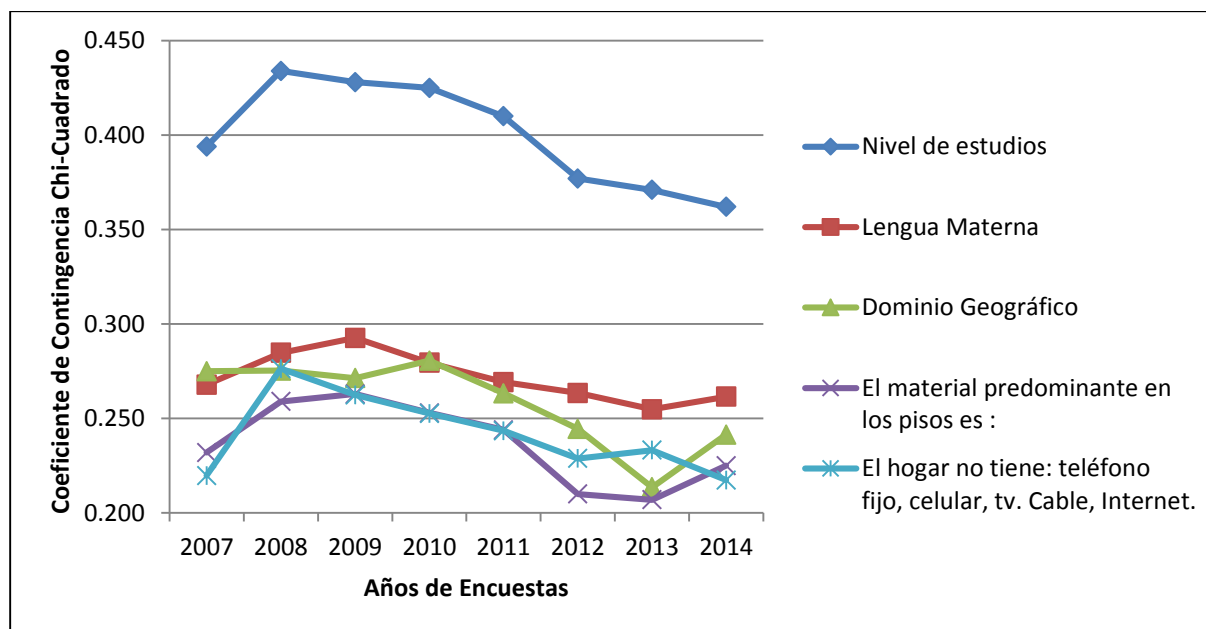
**Ilustración 23:** Combustible que usan en el Hogar para cocinar y la Confianza en el Congreso de la República, 2007, 2013 y 2014

### 3.5.2. Confianza en el Gobierno Regional

En la tabla 9 se detalla los resultados sobre el comportamiento de influencia medidos con el coeficiente de contingencia chi cuadrado, del 2007 al 2014, de las 5 variables más influyentes en la percepción de la confianza en Gobierno Regional.

**Tabla 9:** Resumen de las variables más influyentes en la confianza sobre el Gobierno Regional

N°	Variables más Influyentes	Coeficiente de Contingencia de Chi cuadrado							
		2007	2008	2009	2010	2011	2012	2013	2014
1	Nivel de estudios	0.394	0.434	0.428	0.425	0.410	0.377	0.371	0.362
2	Lengua Materna	0.268	0.285	0.293	0.280	0.269	0.264	0.255	0.262
3	Dominio Geográfico	0.275	0.275	0.271	0.281	0.263	0.245	0.214	0.242
4	El material predominante en los pisos es :	0.232	0.259	0.263	0.253	0.244	0.210	0.207	0.225
5	El hogar no tiene: teléfono fijo, celular, tv. Cable, Internet.	0.220	0.276	0.263	0.253	0.244	0.229	0.233	0.217



**Ilustración 24:** Evolución del coeficiente de contingencia sobre las variables más influyentes en la confianza sobre el Gobierno Regional, del 2007 al 2014.

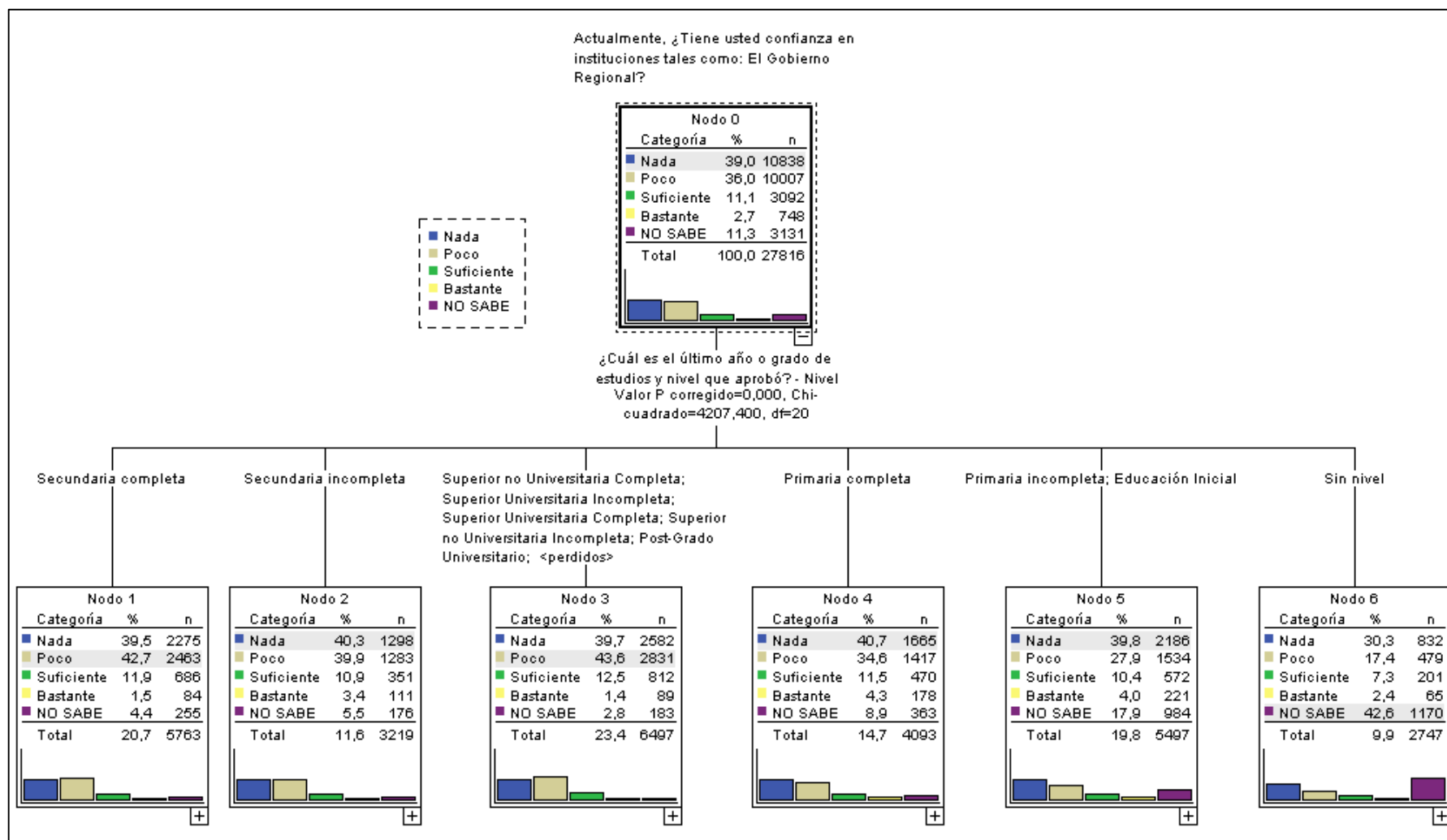
En la tabla 9 se ve las 5 variables más influyentes, el *[Nivel de estudios]* es la primera variable más influyente, desde el 2007 al 2014 y mantiene su posición.

En la ilustración 25, se ve el primer árbol obtenido del algoritmo árbol Chaid, sobre la percepción en la confianza en el Gobierno regional, obteniendo el *[nivel de estudios]* como la variable más influyente y las segmentaciones de sus modalidades de esta variables que proporcionan mayor información, el nivel “superior” tiene alta proporción en relación a “poco de confianza” con un 43,6%, luego “Nada de confianza” de 39,7%. Los porcentajes de “Poco confianza” van disminuyendo cuando él *[nivel de estudios]* baja, también se ve que los de Secundaria incompleta tienen un 39,9%.

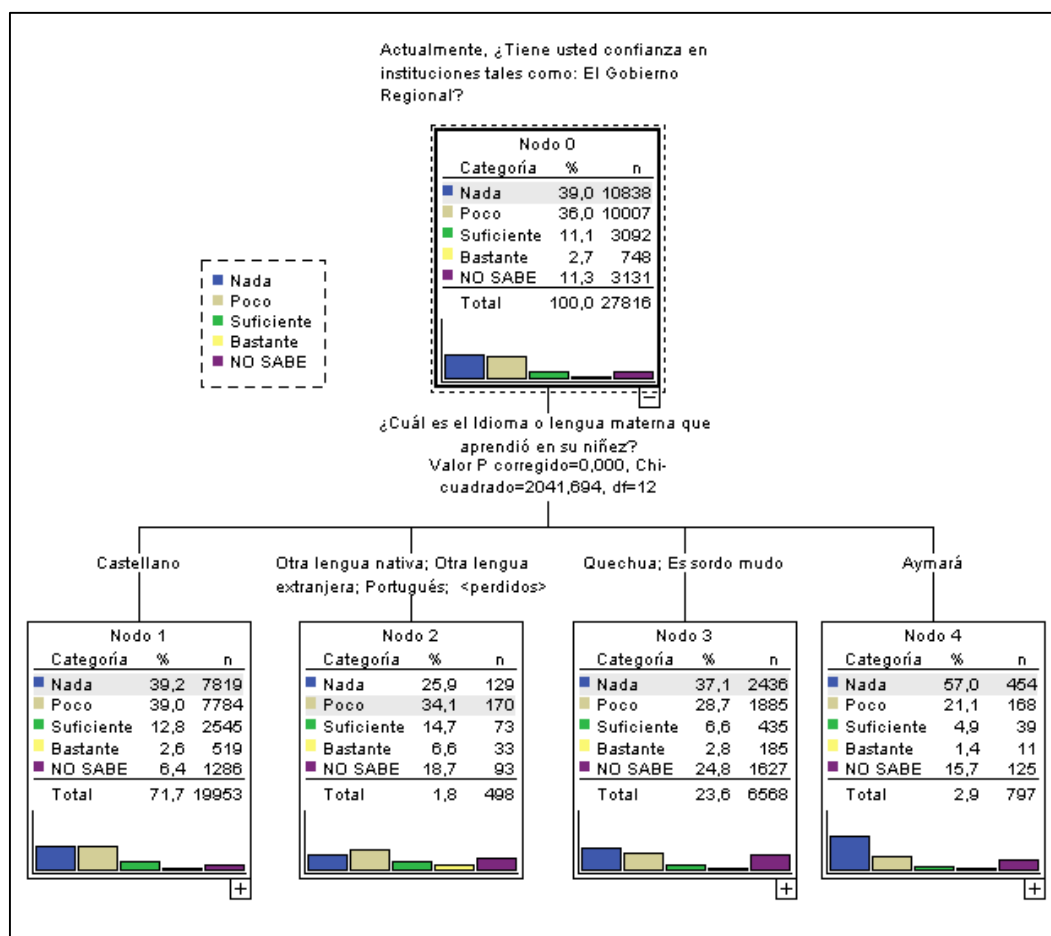
En la ilustración 26, se ve el resultado del segundo árbol chaid, la variable influyente es *[Lengua materna]*, los de habla “Aymara” tienen mayor porcentaje de “Nada de confianza” con un 57%, los de “Quechua” 37,1%, siendo estos segmentos los más altos porcentajes a comparación con “Poco de confianza” y los otros segmentos.

En la ilustración 27, se ve el resultado del árbol chaid en la confianza en el Gobierno Regional, la tercera variable más influyentes es el *[Dominio Geográfico]*, el segmento “sierra” (Norte, Sur y Centro) cuenta con mayor proporción relacionado a “Nada de confianza”, mientras que “Lima metropolitana” con mayor proporción relacionado a “Poco confía”.

En la ilustración 28, se ve el resultado del árbol chaid en la confianza en el Gobierno Regional, la cuarta variable más influyente es el *[Material predominante en los pisos]*, el material de “Cemento” y “tierra” son las cantidades más altas, Cemento (40,6%) y Madera (42,1%) tienen la mayor proporción de “Nada de confianza”, mientras “Tierra” con una proporción más alta que las demás en “No Sabe”.



**Ilustración 25:** Primera variable más Influyente en la Confianza en el Gobierno Regional



**Ilustración 26:** Segunda variable más Influyente en la Confianza en el Gobierno Regional



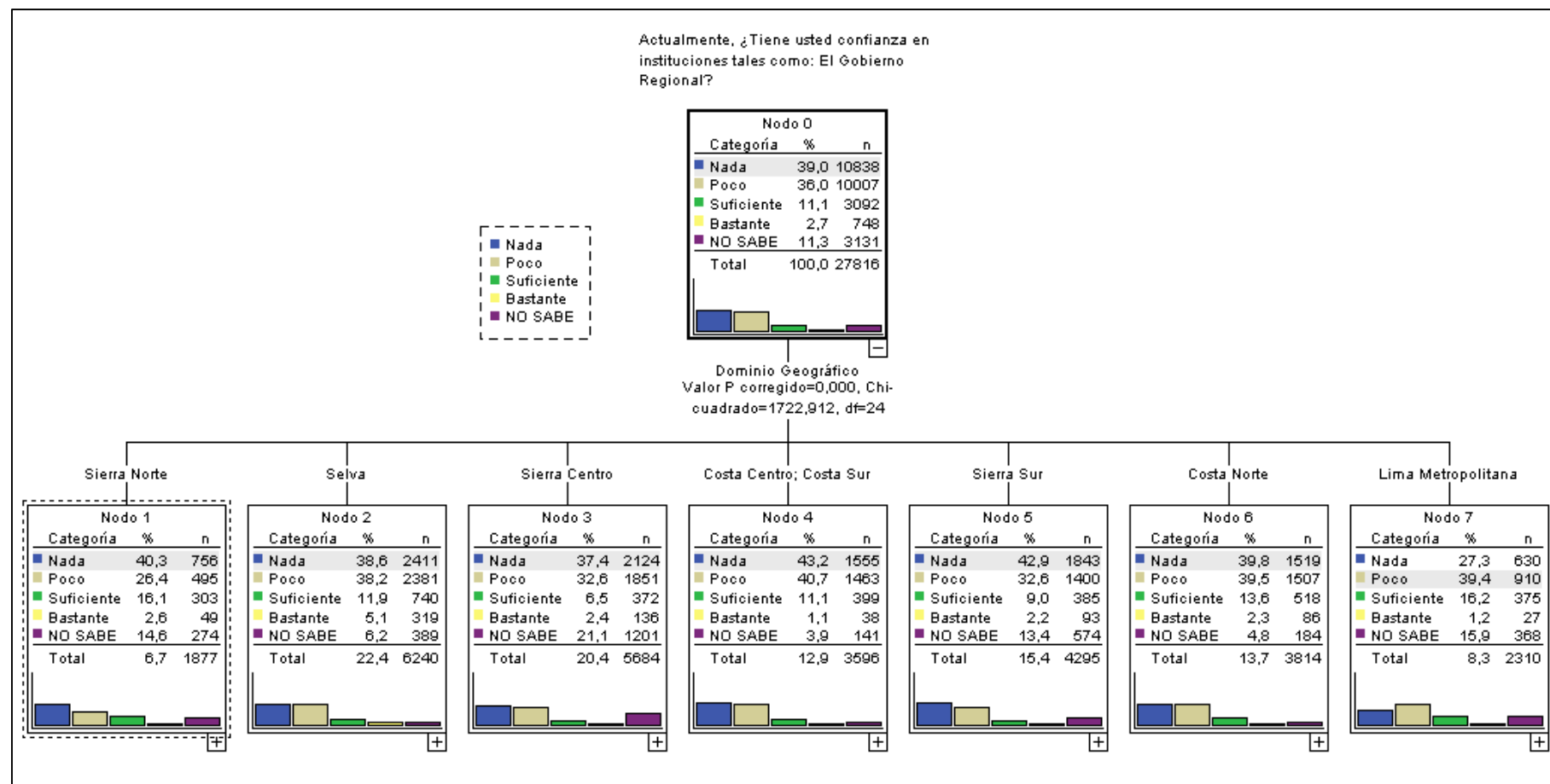


Ilustración 27: Tercera variable más Influyente en la Confianza en el Gobierno Regional

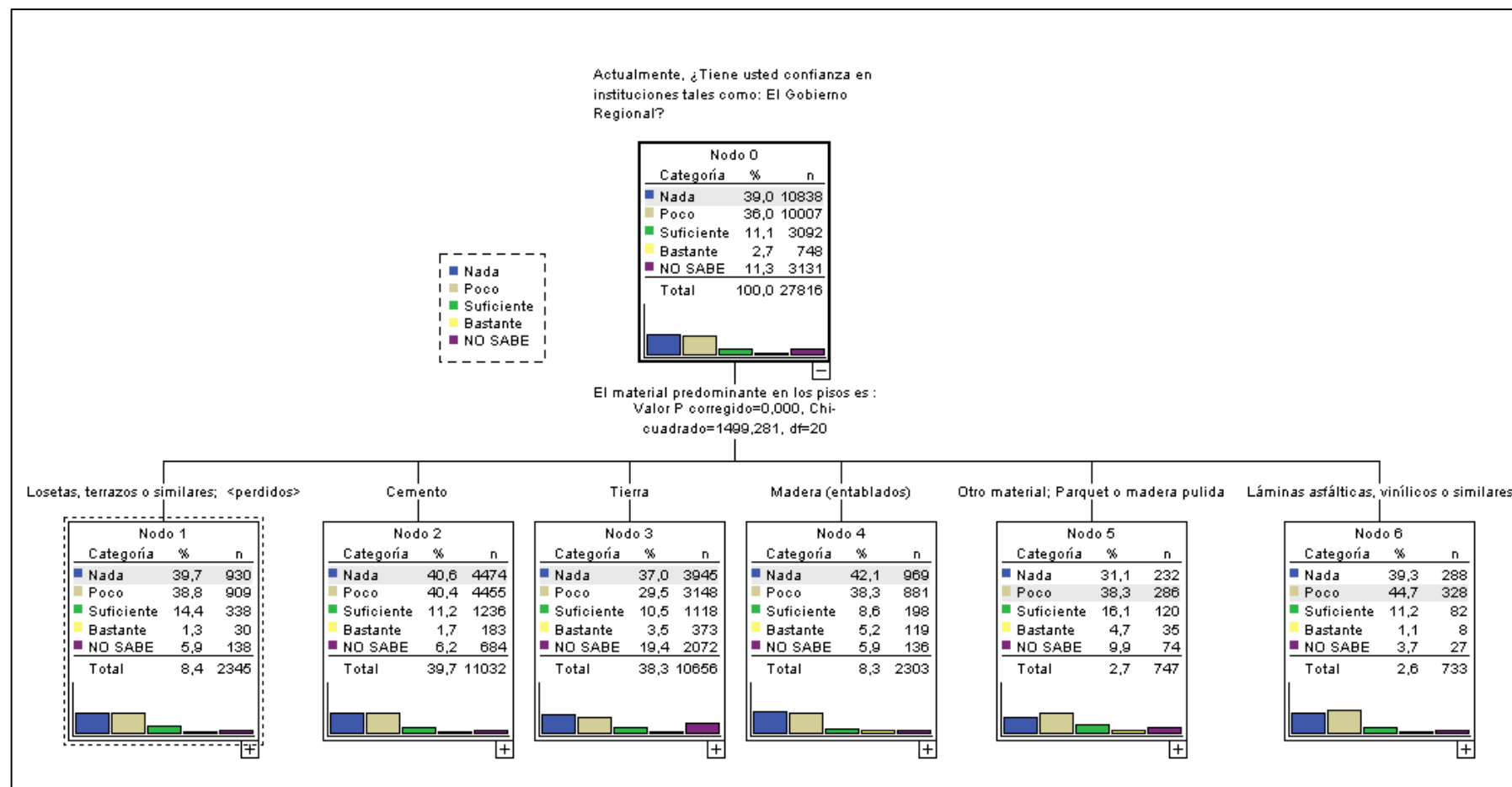
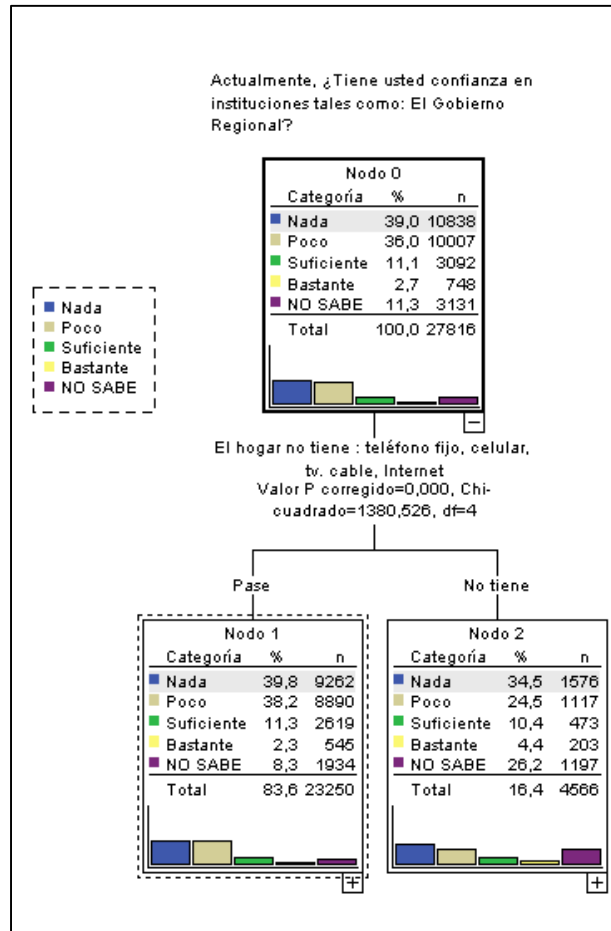


Ilustración 28: Cuarta variable más Influyente en la Confianza en el Gobierno Regional



**Ilustración 29:** Quinta variable más Influyente en la Confianza en el Gobierno Regional

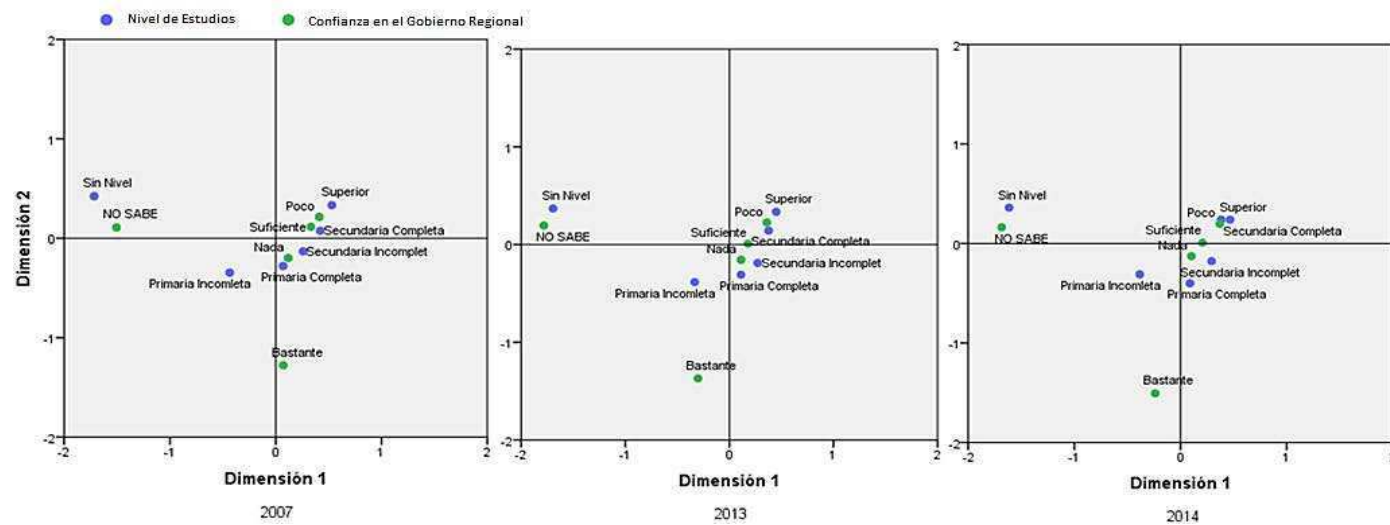
En la ilustración 30 se ven las atracciones entre las modalidades de *[nivel de estudio]* y confianza en el gobierno regional, una comparativa del 2014 con 2007 y 2013. Durante esos años se ven que no hay cambios significativos, “Nada de confianza” está más atraído por nivel de “Primaria completa” y “Secundaria incompleta”, mientras que “Poco” y “Suficiente” confianza esta atraído por Nivel de estudios “Secundaria completa” y “Superior”, “Sin nivel” con “No Sabe”.

En la ilustración 31, se ven las atracciones entre las modalidades de [Lengua materna] y Confianza en el Gobierno Regional, una comparativa del 2014 con 2007 y 2013, 2013 y 2014 se ven sin ningún cambio a comparación del 2007, “Nada de confianza” no tiene una atracción clara, “Poco de confianza” está más relacionado con “Castellano”, también hay una relación muy débil entre “Nada” y lengua Originaria (“Aymara” y “Quechua”), “Quechua” muy atraído con “No Sabe”. El 2007 la relación era más fuerte entre “Nada de confianza” con el segmento de “Aymara” y “Quechua”, “Castellano” con “Poco”, “Suficiente” y “Bastante Confianza”, se podría decir que el 2007 está bien marcado que los que hablan

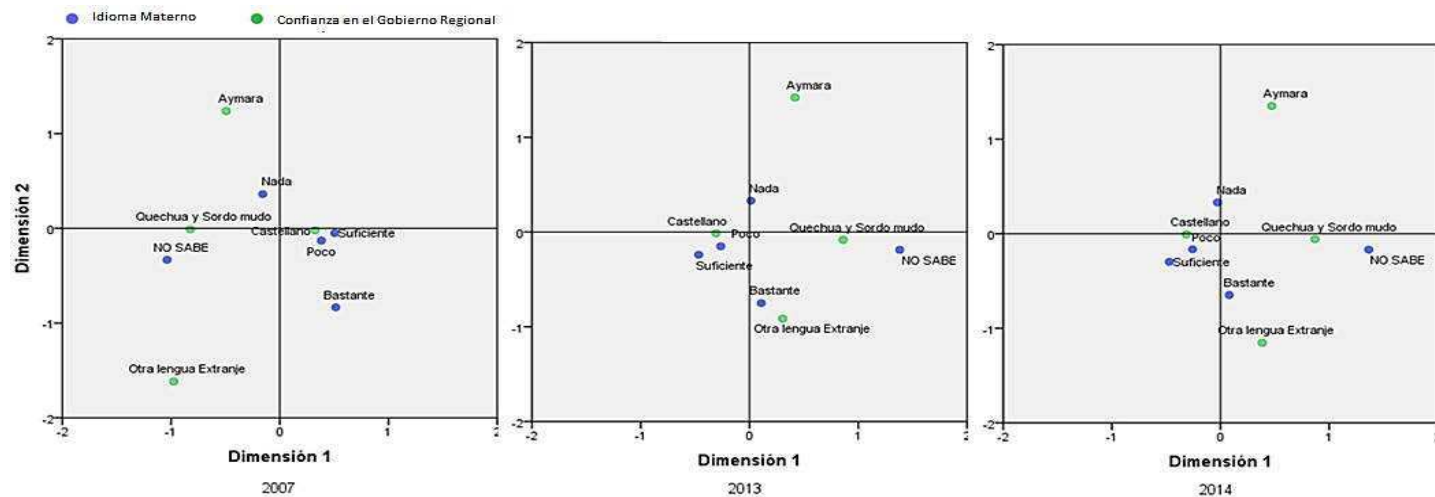
“castellano” tiene más confianza y los que hablan idiomas originarios (“Aymara” y “Quechua”) tiene “Nada” y “No saben” de la confianza.

En la ilustración 32, se ven las atracciones entre [Dominio geográfico] y [Confianza en el Gobierno Regional], se ven variaciones “No sabe” está más atraído de “Sierra centro”, “Sierra Sur” con “Nada de confianza”, Poco y Suficientes en relación con Costa (año 2013).

En la ilustración 33, se ven las atracciones entre *[material predominante en los pisos]* y confianza en el Gobierno Regional, los pisos de “Parquet” o “madera pulida/Otro material” esta atraído por “Poco” y “Nada de confianza”, en el 2014 más atraído por “Nada”, mientras que “Losetas”, “terrazos”, “Laminas asfálticas” esta atraída con “Poco” y “Suficiente confianza”, Tierra muy atraído con No Sabe.



**Ilustración 30:** Nivel de estudios y la Confianza del Gobierno Regional, 2007, 2013 y 2014



**Ilustración 31:** Lengua Materna y la Confianza del Gobierno Regional, 2007, 2013 y 2014

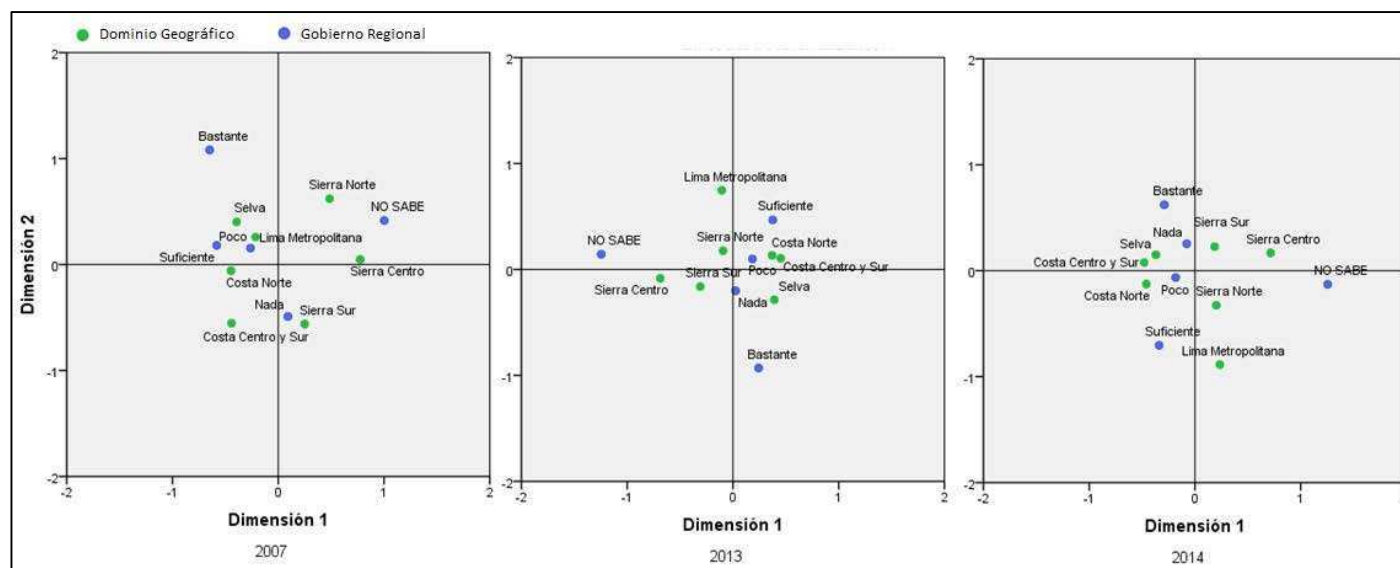


Ilustración 32: Dominio Geográfico y la Confianza del Gobierno Regional, 2007, 2013 y 2014

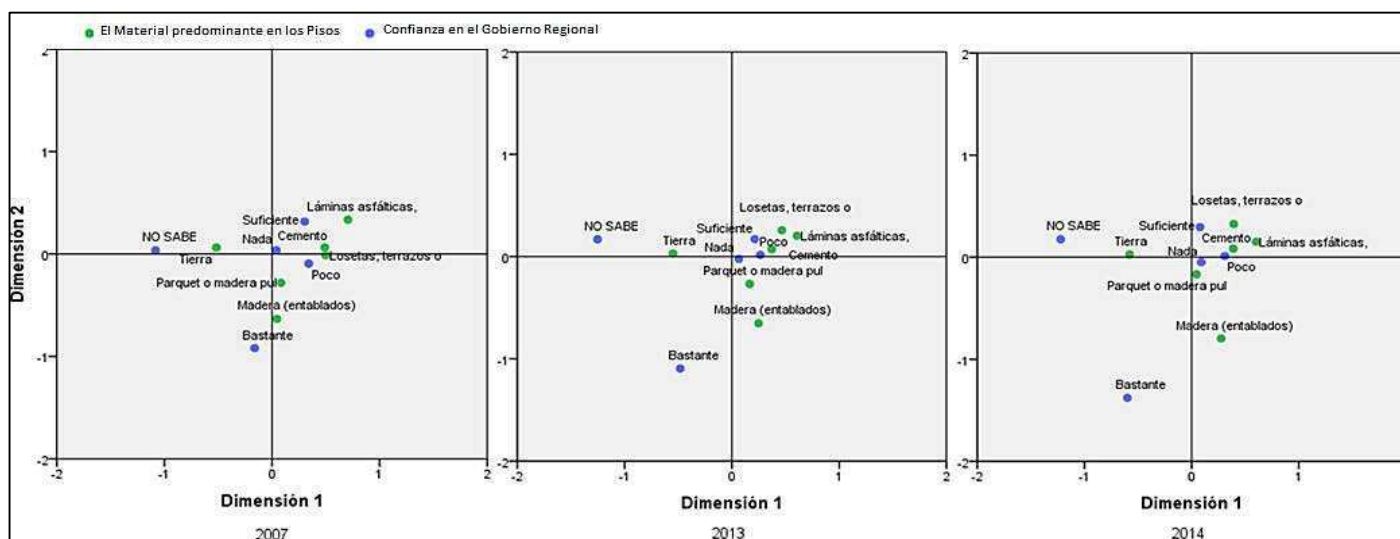
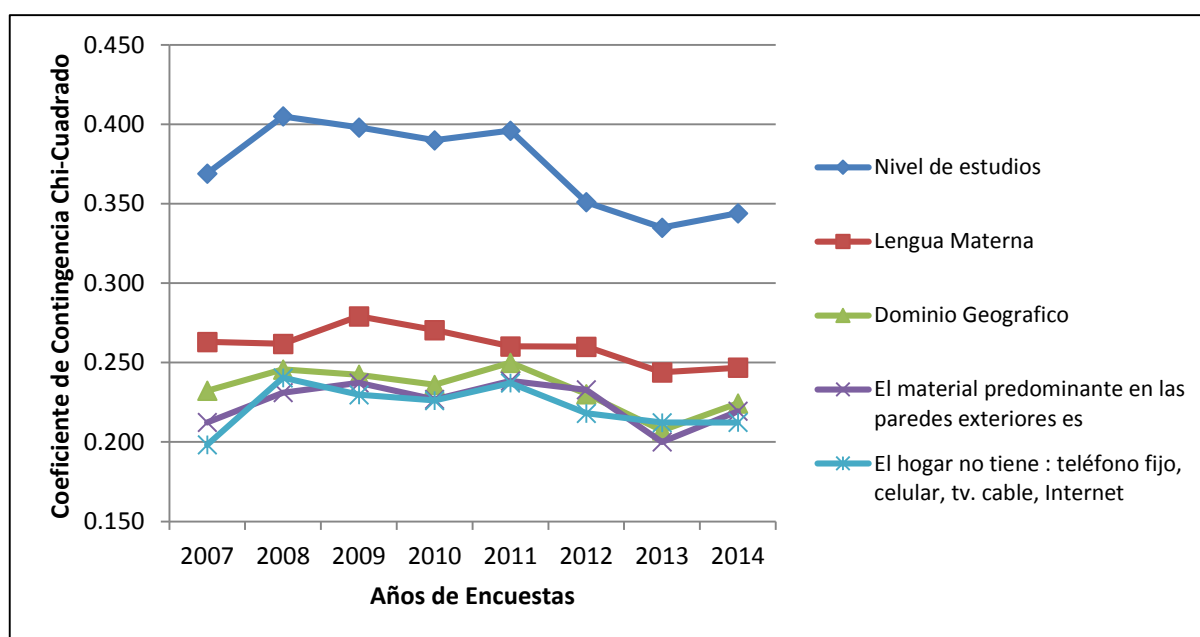


Ilustración 33: El material predominante en los Pisos y la Confianza del Gobierno Regional, 2007, 2013 y 2014

### 3.5.3. Confianza en la Policía Nacional

**Tabla 10:** Resumen de las variables más influyentes en la confianza sobre la Policía Nacional del Perú

N°	Variables más Influyentes	Coeficiente de Contingencia de Chi cuadrado							
		2007	2008	2009	2010	2011	2012	2013	2014
1	Nivel de estudios	0.369	0.405	0.398	0.390	0.396	0.351	0.335	0.344
2	Lengua Materna	0.263	0.262	0.279	0.270	0.260	0.260	0.244	0.247
3	Dominio Geográfico	0.232	0.246	0.242	0.236	0.250	0.230	0.207	0.224
4	El material predominante en las paredes exteriores es.	0.212	0.231	0.237	0.227	0.238	0.233	0.200	0.219
5	El hogar no tiene: teléfono fijo, celular, tv. Cable, Internet.	0.198	0.240	0.230	0.226	0.237	0.218	0.212	0.212



**Ilustración 34:** Evolución del coeficiente de contingencia sobre las variables más influyentes en la confianza sobre la Policía Nacional del Perú, del 2007 al 2014.

En la tabla 10 y la ilustración 34, se ve la evolución de la relación entre las 5 variables más influyentes en la confianza con la policía Nacional, el *[Nivel de estudios]* e *[idioma materno]* permanecen en su posición de influyente del 2007 al 2014, no pasa lo mismo con las 3 variables siguientes, el nivel de influencia general es débil por estar dentro del rango  $<0-0,5]$  (Asociación débil), y se necesita ver a más detalle las atracciones de las modalidades.

En la ilustración 35, se ve el primer árbol obtenido del algoritmo Chaid, sobre la percepción en la confianza en la Policía, obteniendo el *[Nivel de estudios]* como la variable más influyente y las segmentaciones de sus modalidades que proporcionan mayor información en la influencia, los niveles de “Superior no Universitaria Completa” y “Superior universitaria incompleto” con 47,5%, luego “Secundaria Completa” 40,7% y “Superior Universitaria completa y posgrado” 48,8%, “Secundaria Completa y Superior no universitaria incompleta” 43,8% tienen la más alta proporción en cada segmento Poco de confianza, “mientras que Inicial” y “Primaria incompleta” 42,9%, Sin Nivel 39,6% tienen alta proporción en su segmento para “Nada de confianza”, esto puede deducirse que los que están más preparados tienen a tener más confianza que los que no tienen estudios.

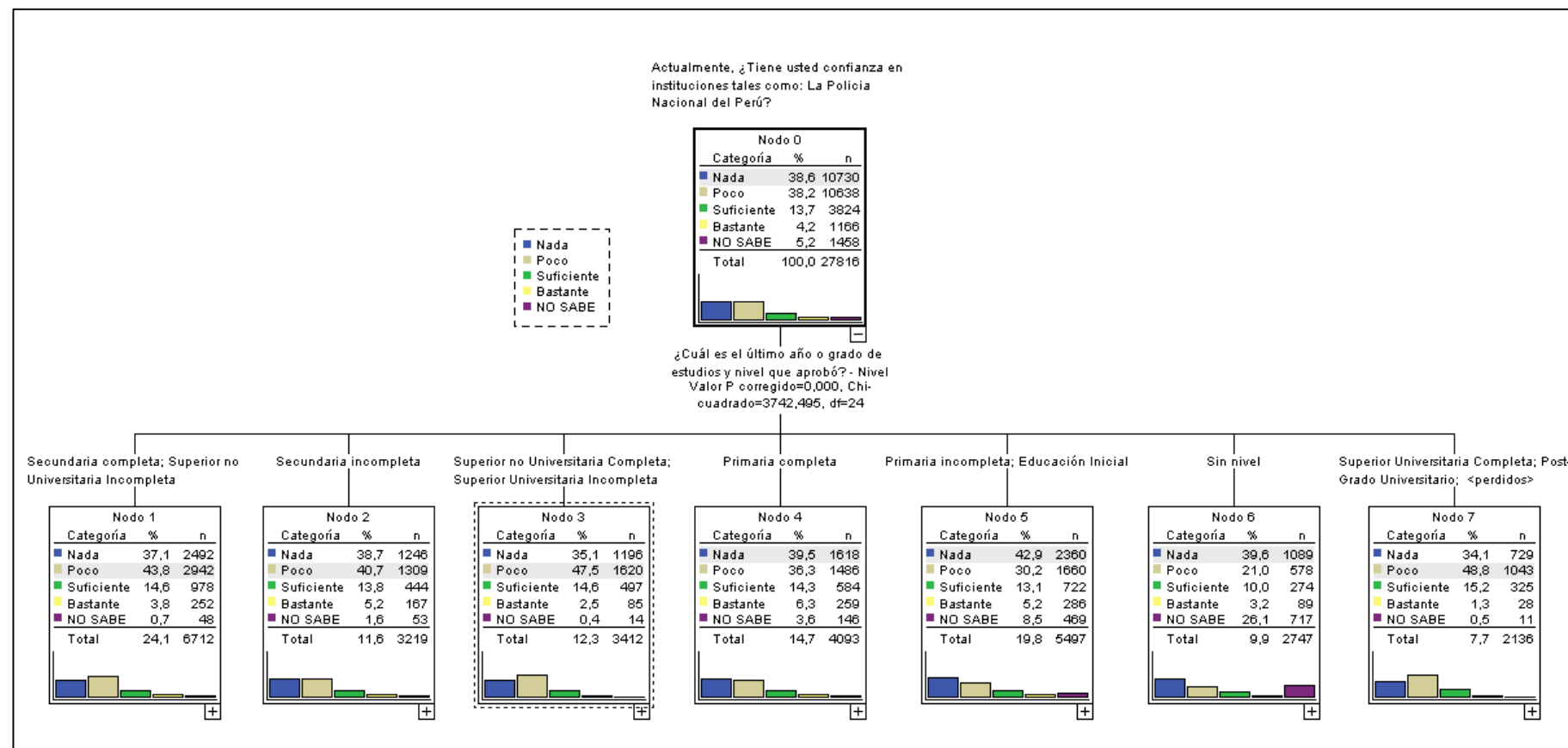
En la ilustración 36, se ve el resultado del segundo árbol chaid, *[Lengua materna]* es la más influyente, la Lengua “Aymara” en un 54,2% y “Quechua/Sordo Mudo” 41,9%, estos representan alto porcentaje relacionado a “Nada de confianza”, como también Lengua “Castellano” con un 41,3% y Otras lenguas Nativas - extranjeras 33,8% tienen alto porcentaje en “Poco de confianza”.

En la ilustración 35, se ve el resultado del tercer árbol chaid, en la confianza en la Policía, la tercera variable más influyentes es el *[Dominio Geográfico]*, la zona de la sierra (“Norte” 41,8%, “Sur” 43,6% y “Centro” 40,8%) tienen alta proporción en “Nada de confianza”, mientras la “Selva” y “Costa” y “Lima metropolitana” alta proporción con respecto a “Poco de confianza”.

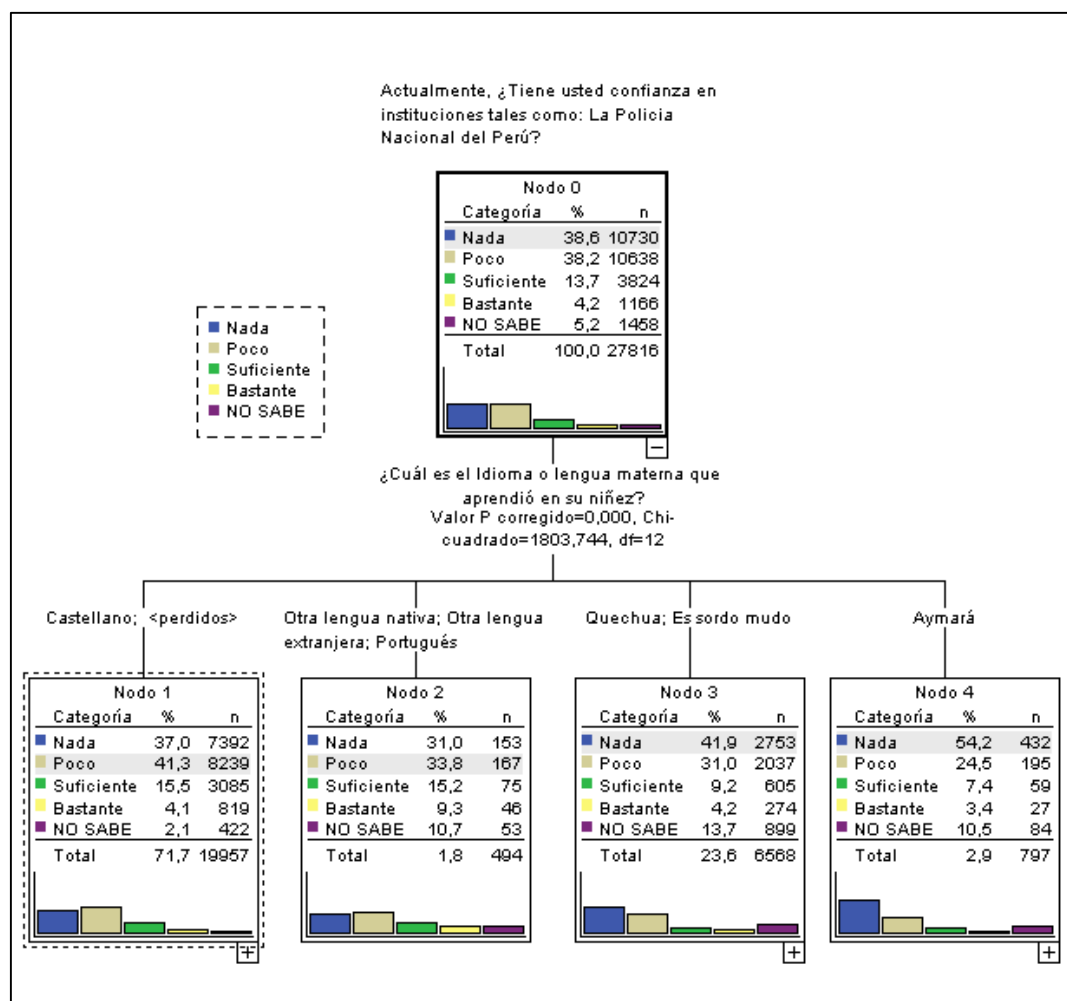
En la ilustración 38, se ve el resultado del cuarto árbol chaid en la confianza en la Policía, la cuarta variable más influyente es el *[Material predominante en las paredes]*, donde el material “Tapia” 44%, “Piedra con Barro” 43,1% y “Adobe/Piedra o Sillar 39,2%, son los más altos porcentajes en “Nada de confianza”, luego también “Ladrillo” o “bloque de cemento” 43,2%, “Otro material” 37,2% y “Madera o quinchá” 39,7% son altos porcentajes en “Poco de confianza”.

En la ilustración 39, se ve el resultado del quinto árbol chaid en la confianza en la Policía, la Quinta variable más influyente es *[El hogar no tiene: teléfono fijo, celular, tv, cable, Internet]*, siendo “No tiene” en un 38% con respecto a “Nada de confianza” siendo la mayor proporción del segmento y 15,5% “No Sabe”, el segmento de “Tiene el servicio” con respecto de “Poco de confianza” con 40,4%, como también los que “No tienen estos servicios” tienen a desconfiar .

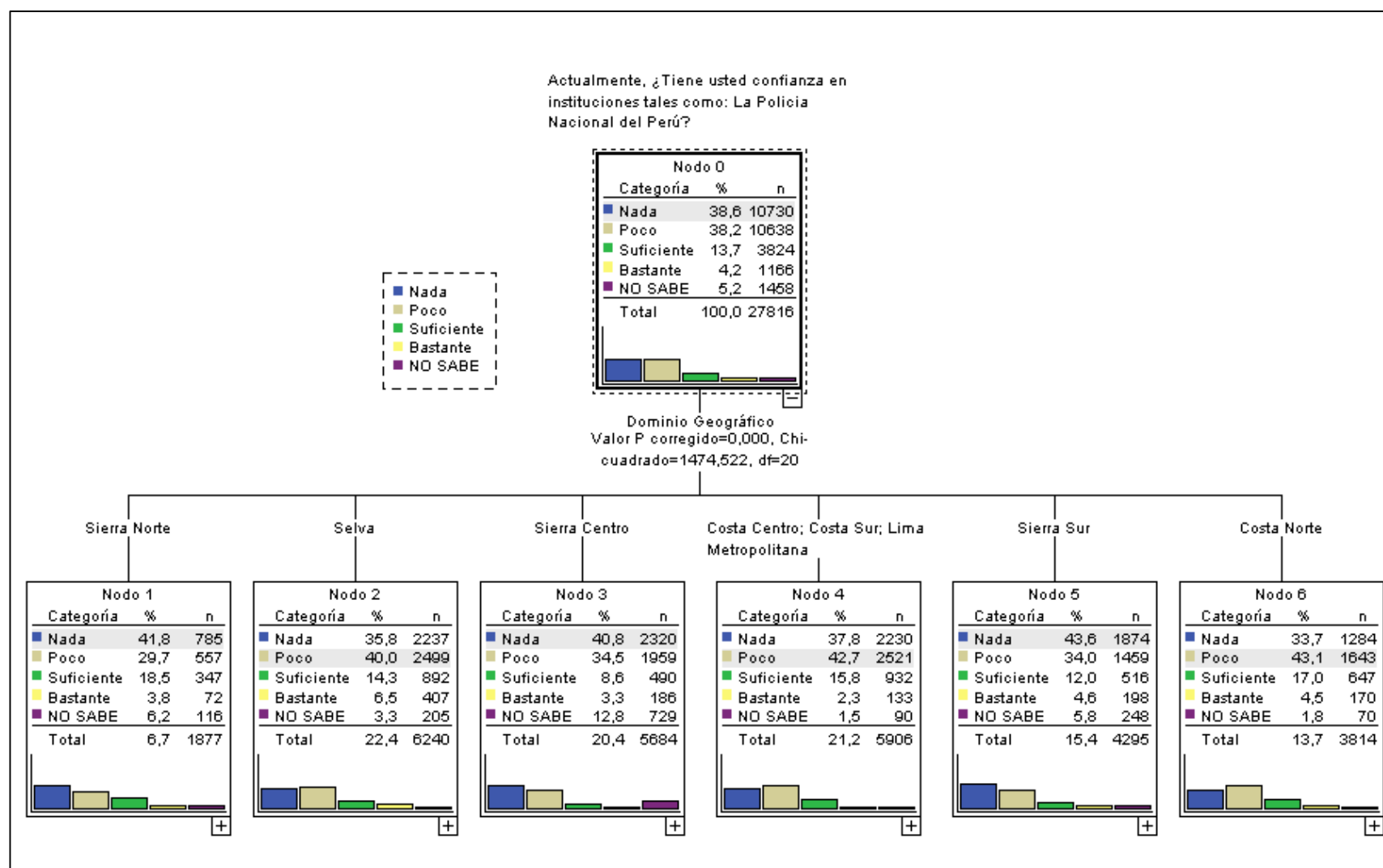




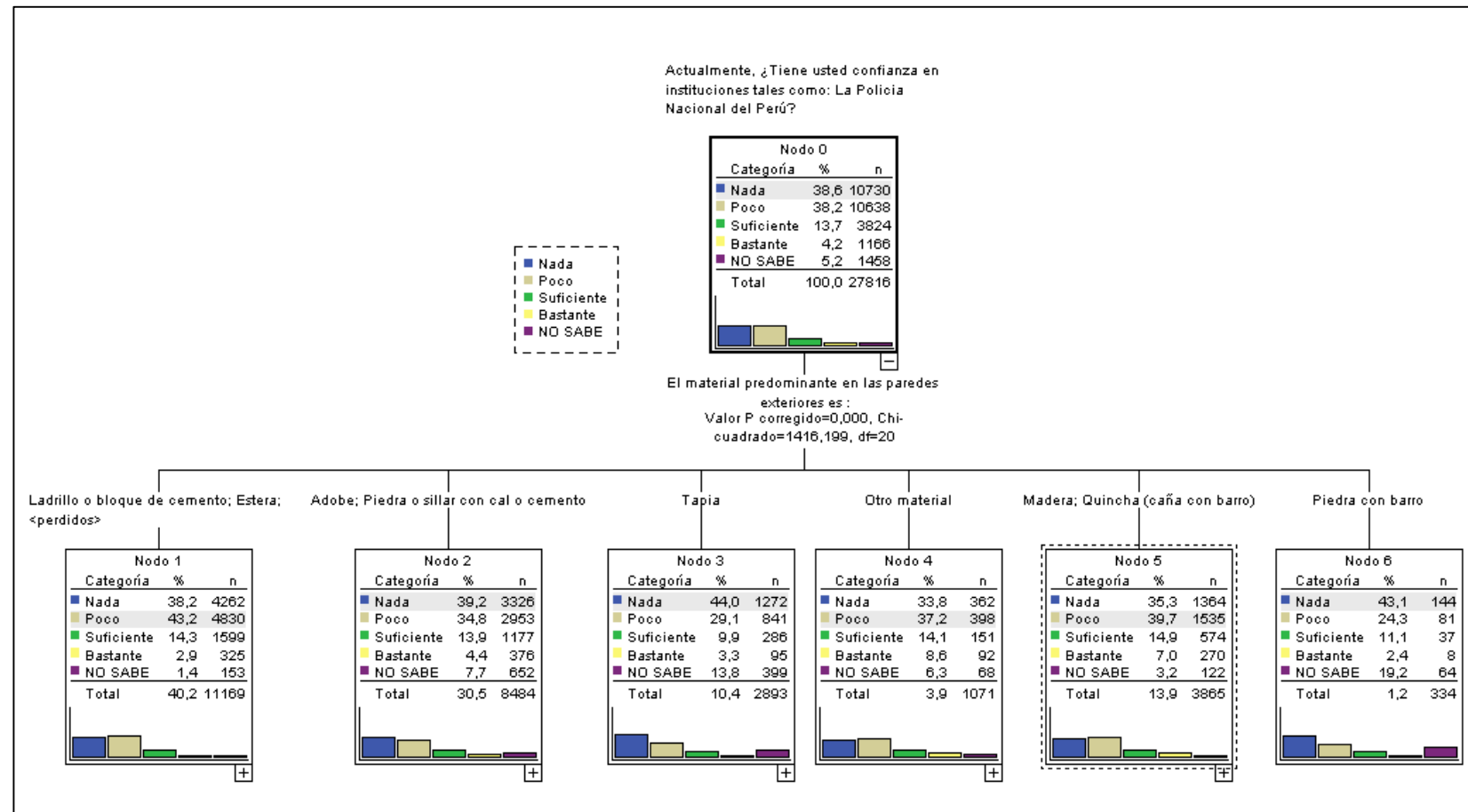
**Ilustración 35:** Primera variable más Influyente en la Confianza en la Policía Nacional



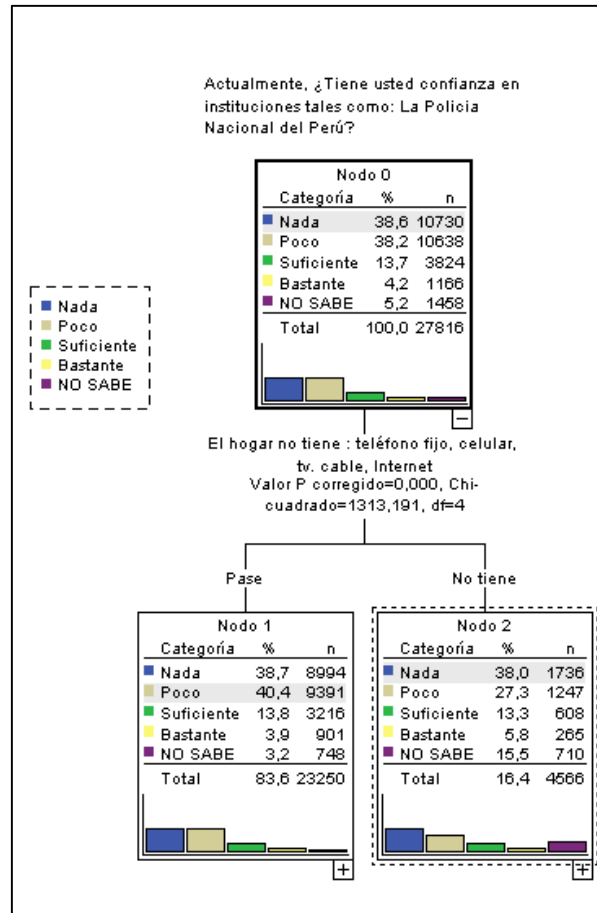
**Ilustración 36:** Segunda variable más Influyente en la Confianza en la Policía Nacional



**Ilustración 37:** Tercera variable más Influyente en la Confianza en la Policía Nacional



**Ilustración 38:** Cuarta variable más Influyente en la Confianza en la Policía Nacional



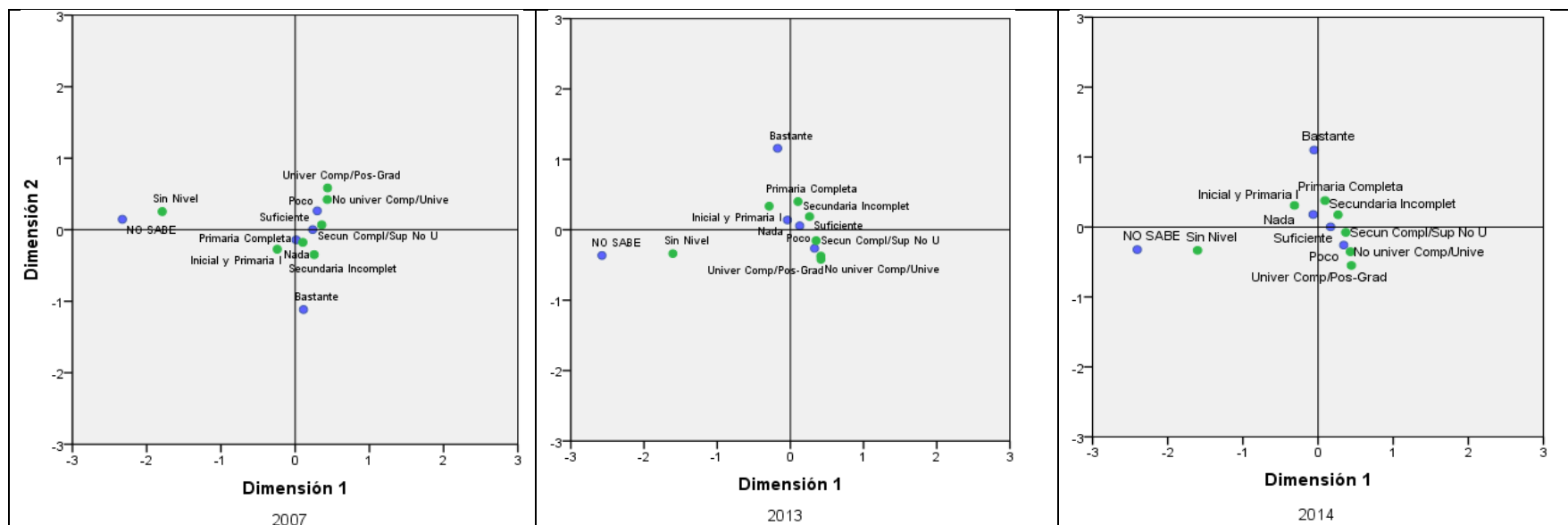
**Ilustración 39:** Quinta variable más Influyente en la Confianza en la Policía Nacional

En la ilustración 40 se ven las atracciones entre las modalidades de *[nivel de estudio]* y confianza en la Policía, una comparativa del 2014 con 2007 y 2013. Niveles alto de educación están más atraídos a “Poco de confianza”, el nivel “Inicial” y “Primaria” están más atraídos por “Nada de Confianza”.

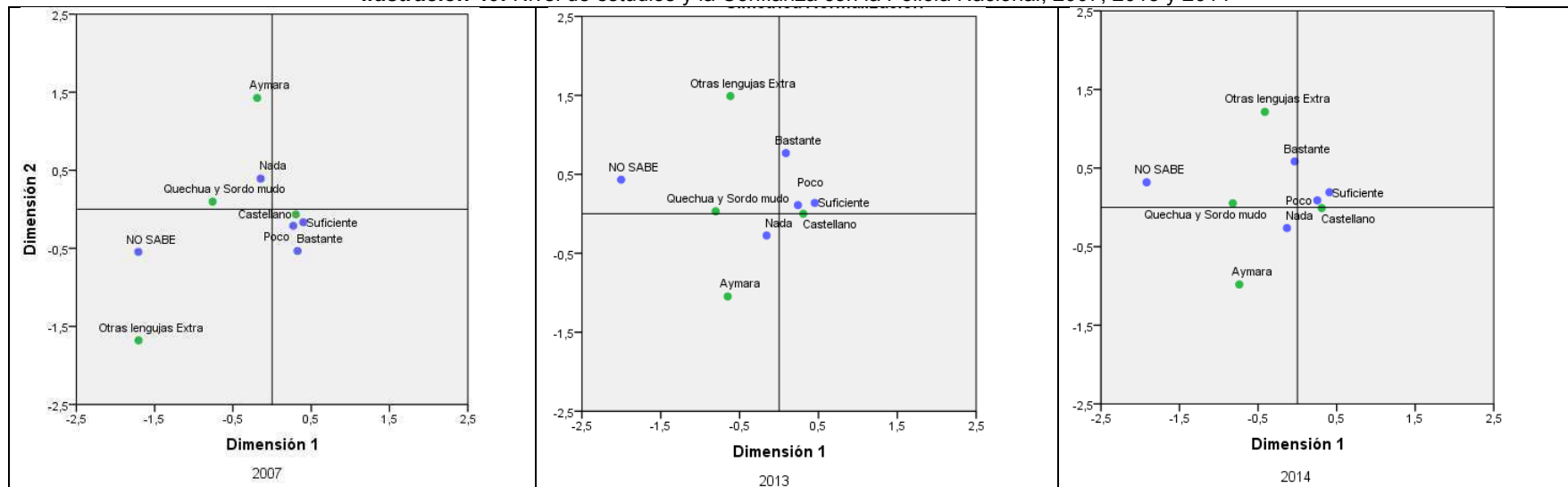
En la ilustración 41, se ven las atracciones entre las modalidades de *[Lengua Materna]* y Confianza en la policía, una comparativa del 2014 con 2007 y 2013, se ve que hablantes de “castellano” son más atraídos por tener confianza, y de Lenguas originarios como él (“Aymara” y “quechua”) más atraídos a “Nada de confianza” y “No sabe”.

En la ilustración 42, se ven las atracciones entre *[Dominio geográfico]* y Confianza en la policía, la “sierra” está más atraída por “Nada y No sabe” de confianza, mientras que la “costa” y “selva” más atraído a “Poco”, “Suficiente” y “Bastante” confianza.

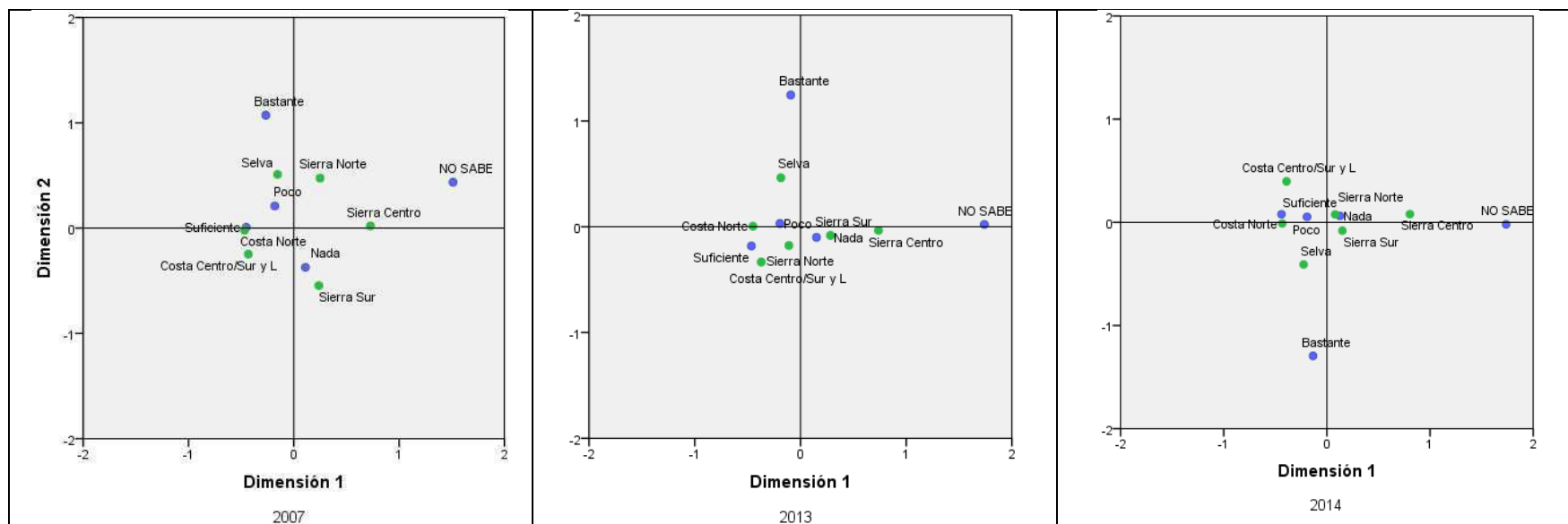
En la ilustración 43, se ven las atracciones entre el *[material predominante en los pisos]* y confianza en la Policía, los pisos de “Piedra o sillar” están más atraídos por “Nada” de confianza, “Ladrillo” o “bloque” con “Poco” y “Suficiente” confianza, como también “Adobe” y “quincha” más atraído con “No sabe”.



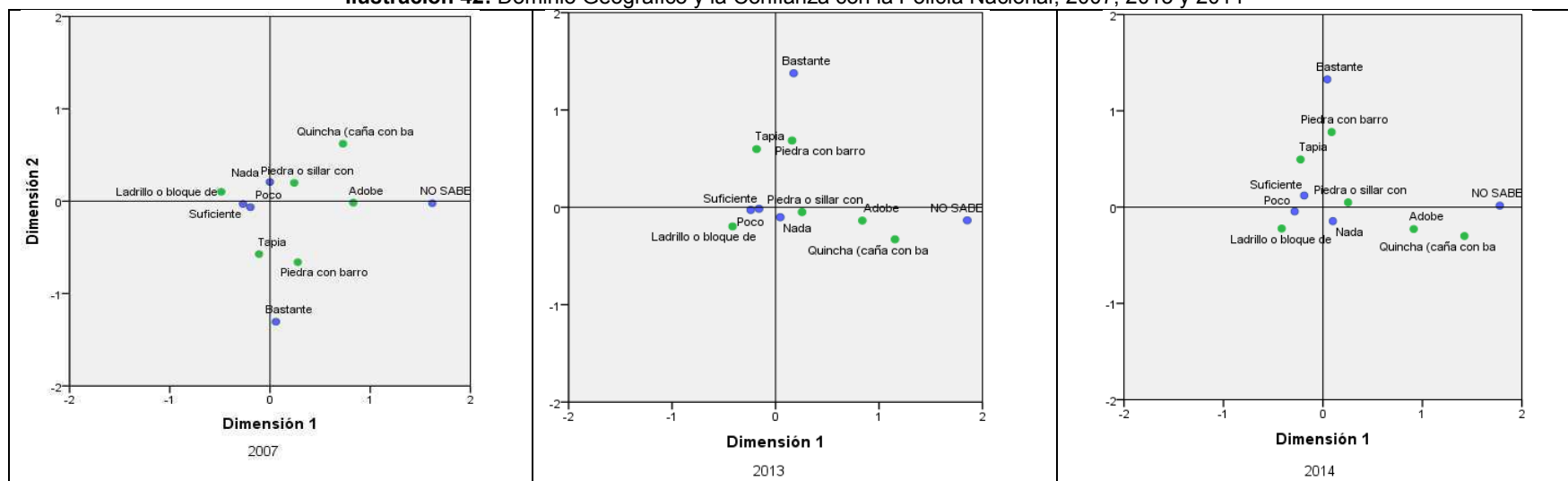
**Ilustración 40:** Nivel de estudios y la Confianza con la Policía Nacional, 2007, 2013 y 2014



**Ilustración 41:** Lengua Materna y la Confianza con la Policía Nacional, 2007, 2013 y 2014



**Ilustración 42:** Dominio Geográfico y la Confianza con la Policía Nacional, 2007, 2013 y 2014

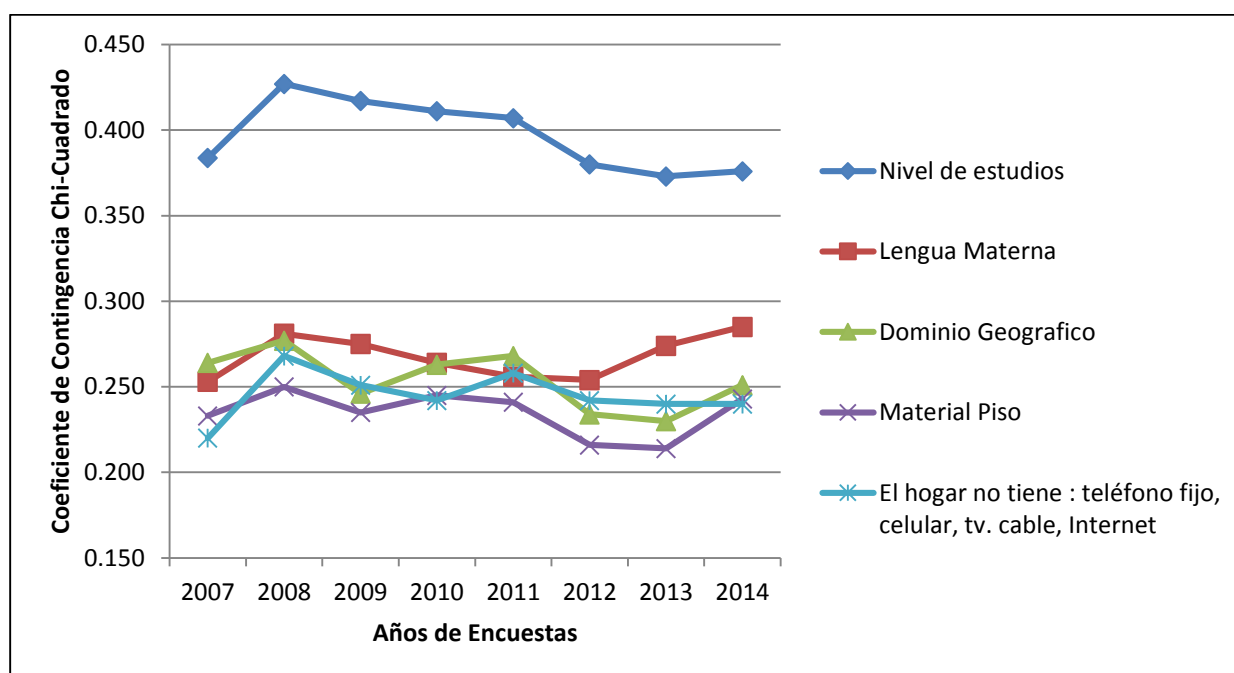


**Ilustración 43:** Material Predominante en las Paredes y la Confianza del Policía Nacional, 2007, 2013 y 2014

### 3.5.4. Confianza en el Poder Judicial

**Tabla 11:** Resumen de las variables más influyentes en la confianza sobre el Poder Judicial

N°	Variables más Influyentes	Coeficiente de Contingencia de Chi cuadrado							
		2007	2008	2009	2010	2011	2012	2013	2014
1	Nivel de estudios	0.384	0.427	0.417	0.411	0.407	0.380	0.373	0.376
2	Lengua Materna	0.253	0.281	0.275	0.264	0.256	0.254	0.274	0.285
3	Dominio Geográfico	0.264	0.277	0.246	0.263	0.268	0.234	0.230	0.251
4	Material Piso	0.233	0.250	0.235	0.245	0.241	0.216	0.214	0.243
5	El hogar no tiene: teléfono fijo, celular, tv. Cable, Internet.	0.220	0.268	0.251	0.242	0.258	0.242	0.240	0.240



**Ilustración 44:** Evolución del coeficiente de contingencia sobre las variables más influyentes en la confianza sobre el Poder Judicial, del 2007 al 2014.

En la tabla 11 y la ilustración 44, se ve la evolución de la relación entre las 5 variables más influyentes en la confianza con el Poder Judicial, el *[Nivel de estudios]* permanece en su ubicación influyente durante el 2007 al 2014, no pasa lo mismo con las 4 variables siguientes, el nivel de influencia general es débil está dentro del rango  $<0-0,5]$  (Asociación débil), y se necesita ver a más detalle las atracciones de las modalidades.

En la ilustración 45, se ve el primer árbol obtenido del algoritmo Chaid, sobre la percepción en la confianza en el Poder Judicial, los segmentos distribuidos “Primaria



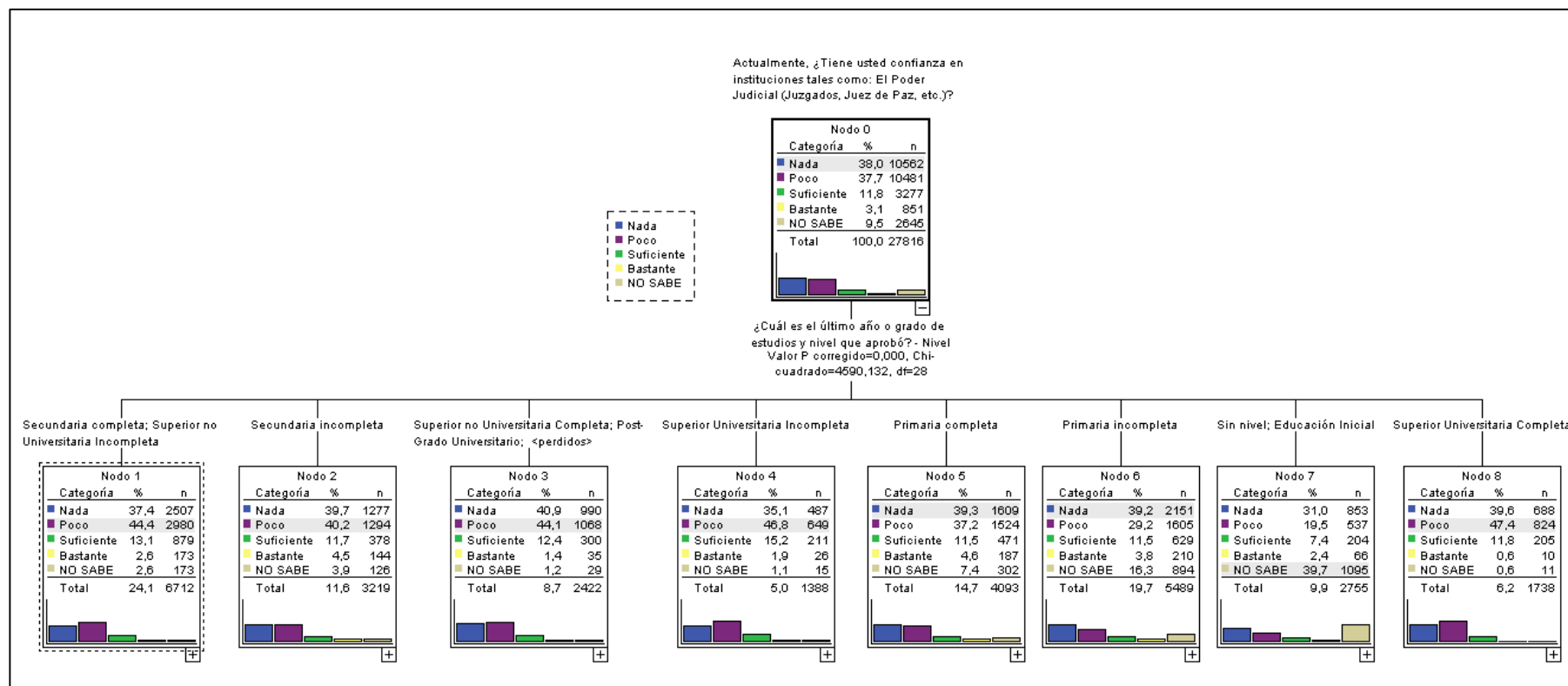
incompleta” son del 39,3%, “Primaria completa” del 39,2% y “Sin nivel” y “Primaria” del 31% estos son los que muestran alta proporción en “Nada de confianza”, mientras que “Superior Universitaria completa” con 47,4%, “Superior Universitaria incompleta” con 46,8, “secundaria completa” y “Superior no universitaria incompleta” con 44,4% relacionado con “poco de confianza”.

En la ilustración 46, se ven el resultado del segundo árbol chaid, sobre la percepción en la confianza en el Poder Judicial, la variable influyente es [*Legua materna*], en los segmentos los de habla “Aymara” con 55,8% y “Quechua”/”Sordo Mudo” con 37,1%, estos con alta proporción relacionado en “Nada de confianza”, mientras que “Castellano y Portugués” con un 40,8% y “Otras lenguas Nativas” con 30,4%, alta proporción relacionado con “Poco de confianza”.

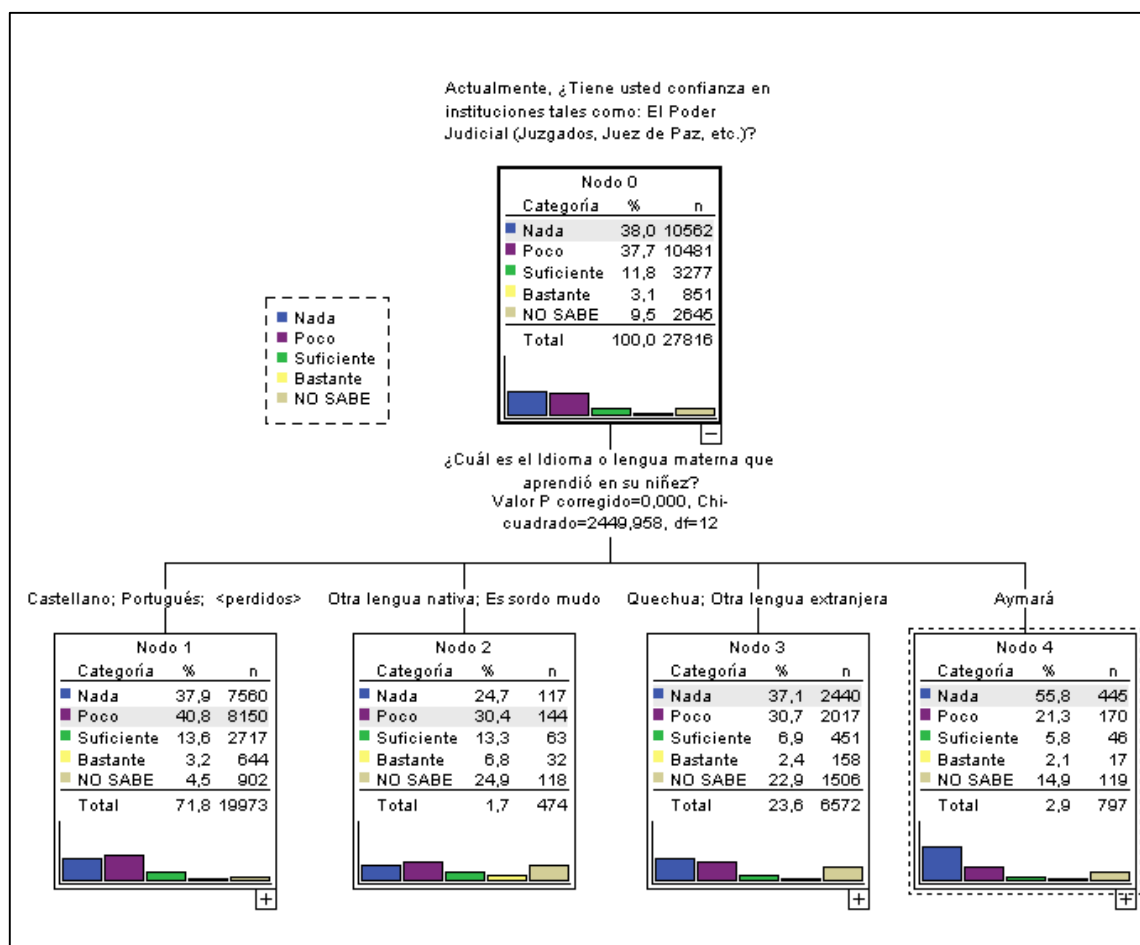
En la ilustración 47, se ve el resultado del árbol chaid en la confianza en el poder judicial, la tercera variable más influyentes es el [*Dominio Geográfico*], donde la “sierra” y “Centro” tienen alta proporción relacionada a “Nada de confianza”, mientras la “Selva”, “Costa” y “Lima metropolitana” con alta proporción en “Poco confía”.

En la ilustración 48, se ve el resultado del árbol chaid en la confianza en el Poder Judicial, la cuarta variable más influyente es el [*Material predominante en el Piso*], el valor de “Tierra y otro material” es el de mayor proporción relacionada a “Nada de confianza”, mientras que “Cemento” tiene alta proporción “Poco de confianza”.

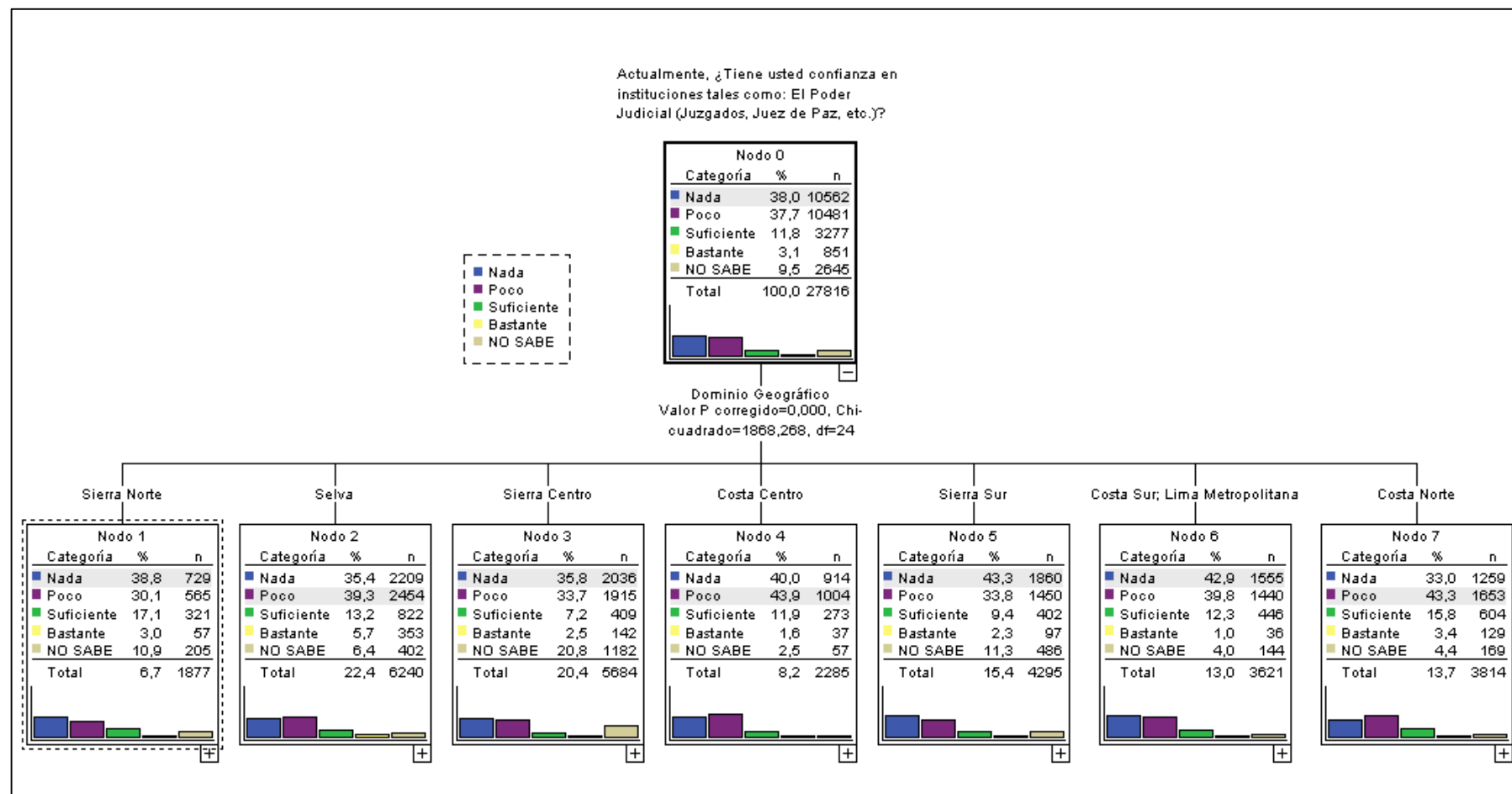
En la ilustración 49, se ve el resultado del árbol chaid en la confianza en la Policía, la Quinta variable más influyente es [El hogar no tiene: teléfono fijo, celular, tv, cable, Internet], con “No tiene” con 33,6% con alta proporción relacionada a “Nada de confianza”, y 25,3% relacionado a “No Sabe”, mientras que el segmento de “Tiene” con 40% relacionado con “Poco de confianza”, los que No tienen estos servicios tienen a desconfiar.



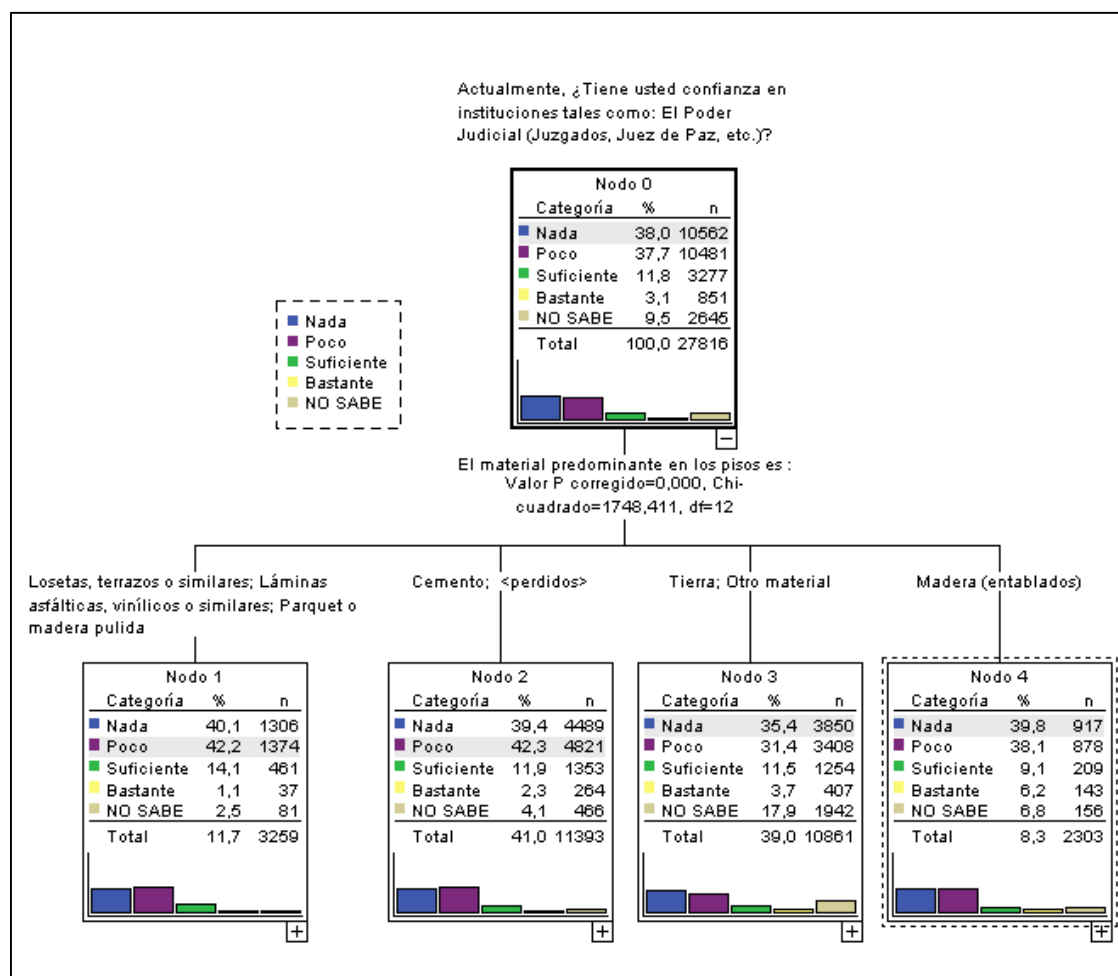
**Ilustración 45:** Primera variable más Influyente en la Confianza en el Poder Judicial



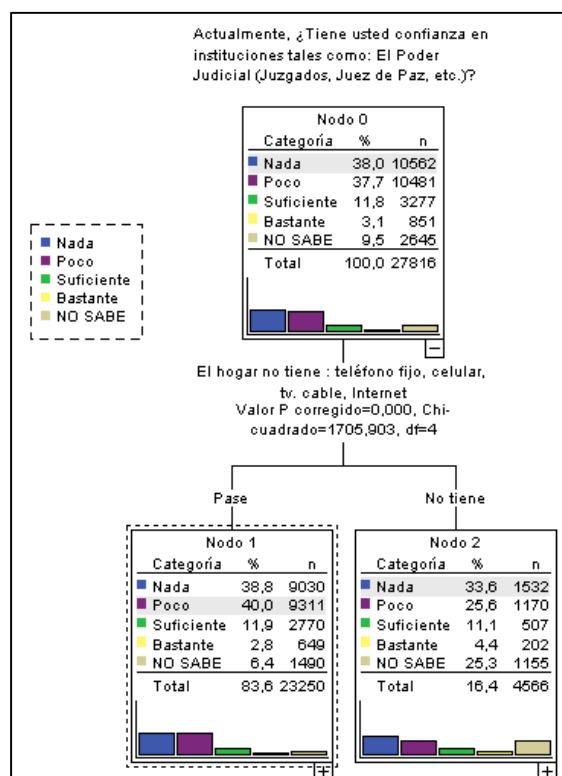
**Ilustración 46:** Segunda variable más Influyente en la Confianza en el Poder Judicial



**Ilustración 47:** Tercera variable más Influyente en la Confianza en el Poder Judicial



**Ilustración 48:** Cuarta variable más Influyente en la Confianza en el Poder Judicial



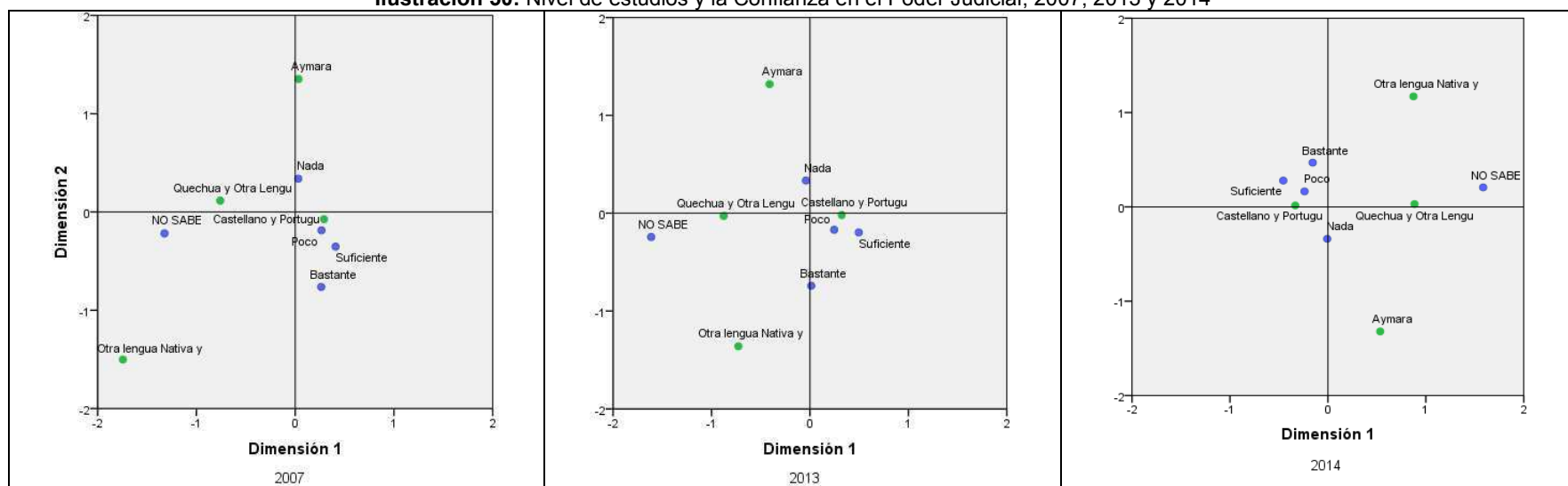
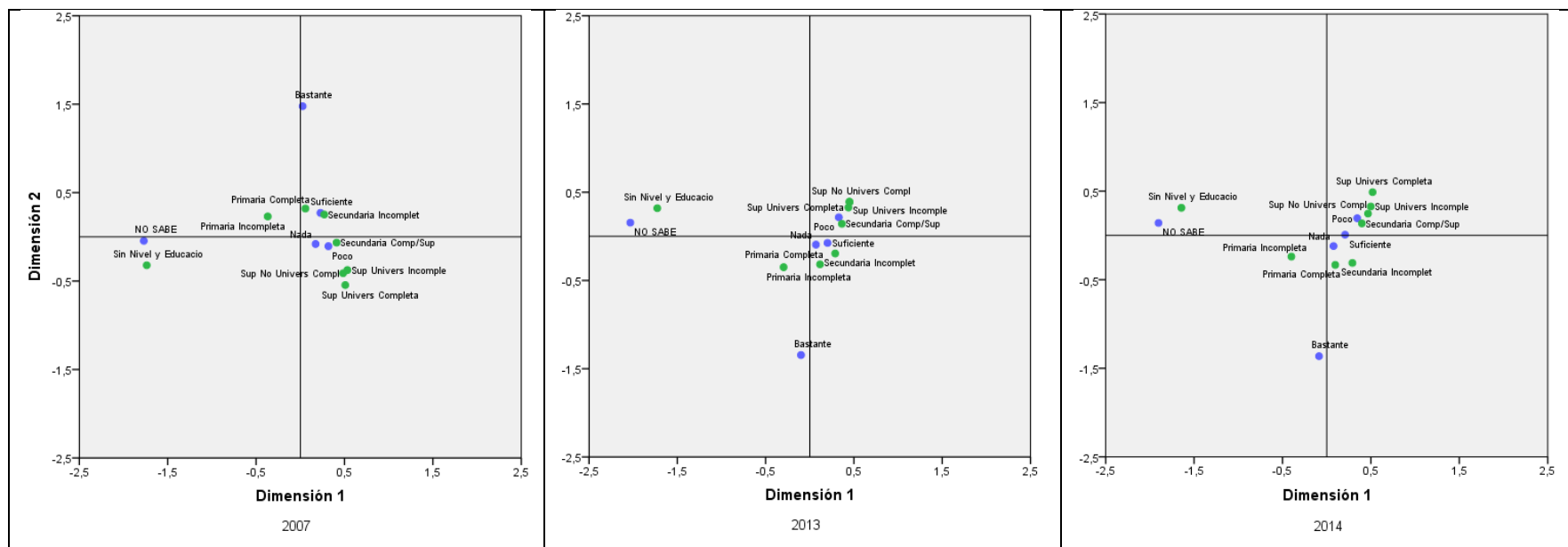
**Ilustración 49:** Quinta variable más Influyente en la Confianza en el Poder Judicial

En la ilustración 50 se ven las atracciones entre las modalidades de *[nivel de estudio]* y confianza en el Poder Judicial, una comparativa del 2014 con 2007 y 2013. Durante esos años se ven que no hay cambios significativos, “Nada de confianza” está más atraído por nivel de “Primaria completa” y “Secundaria incompleta”, mientras que “Poco” y “Suficiente” de confianza esta atraído por “Secundaria completa” y “Superior”, mientras que “Sin nivel” es atraído con “No Sabe”.

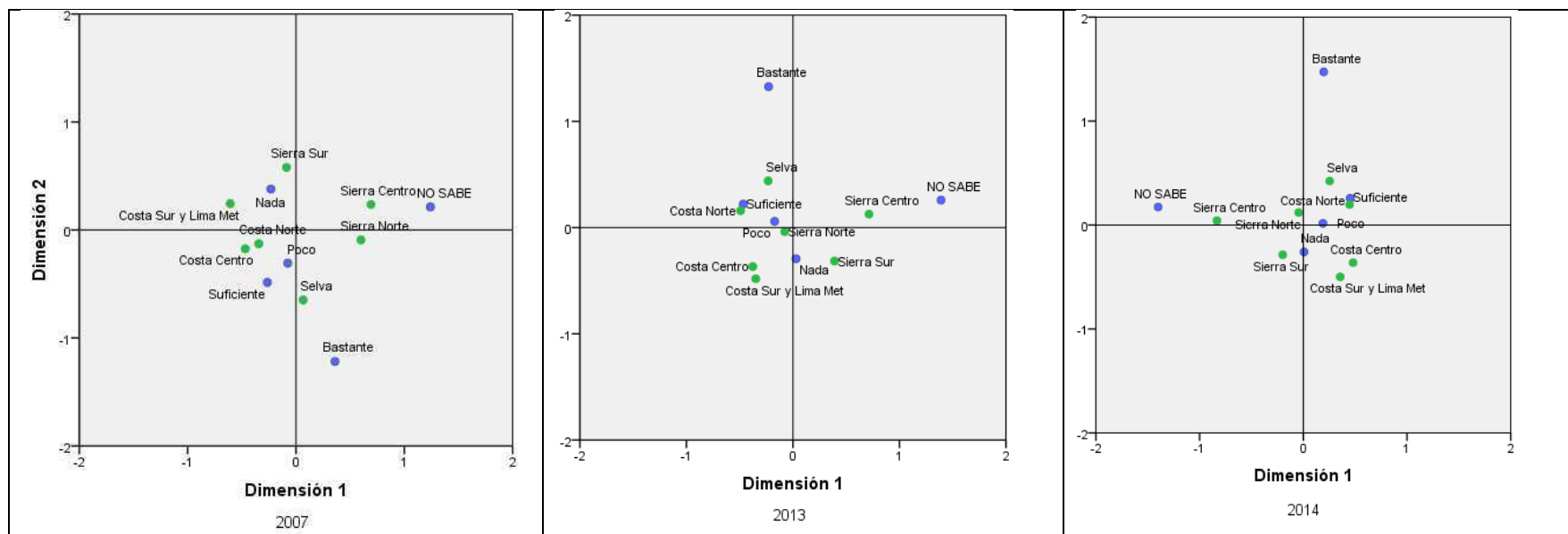
En la ilustración 51, se ven las atracciones entre las modalidades de [Lengua materna] y Confianza en el Poder Judicial, una comparativa del 2014 con 2007 y 2013, “Nada” de confianza no tiene una atracción clara, mientras que “Poco”, “Suficiente” y “Bastante” de confianza está más relacionado con “Castellano”, también hay una relación muy débil entre “Nada” y lengua Originaria (“Aymara” Y “Quechua”), “Quechua” muy atraído con No Sabe.

En la ilustración 52, se ven las atracciones entre *[Dominio geográfico]* y Confianza en el Gobierno Regional, “Sierra sur” es atraída por “Nada de confianza”, “Costa Norte” con “Suficiente Confianza”, mientras “No sabe” con “Sierra centro”.

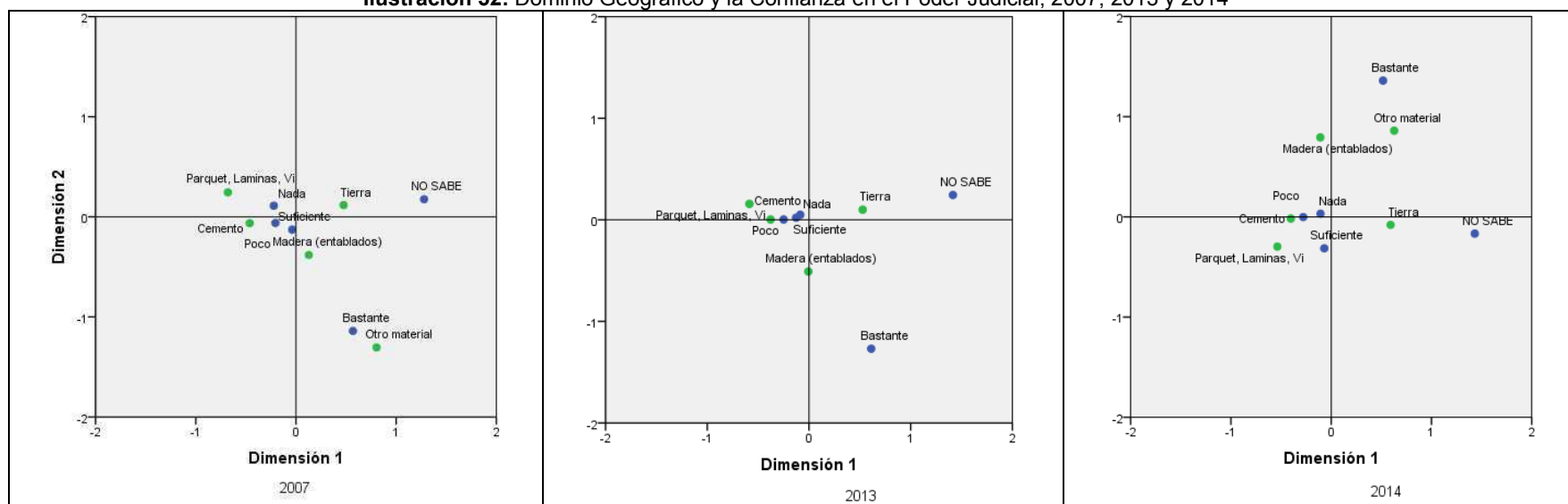
En la ilustración 53, se ven las atracciones entre [material predominante en los pisos] y confianza en el Poder Judicial, los pisos de “Tierra” están más asociados a “No Saben”, “Cemento” con “Poco de confianza”.





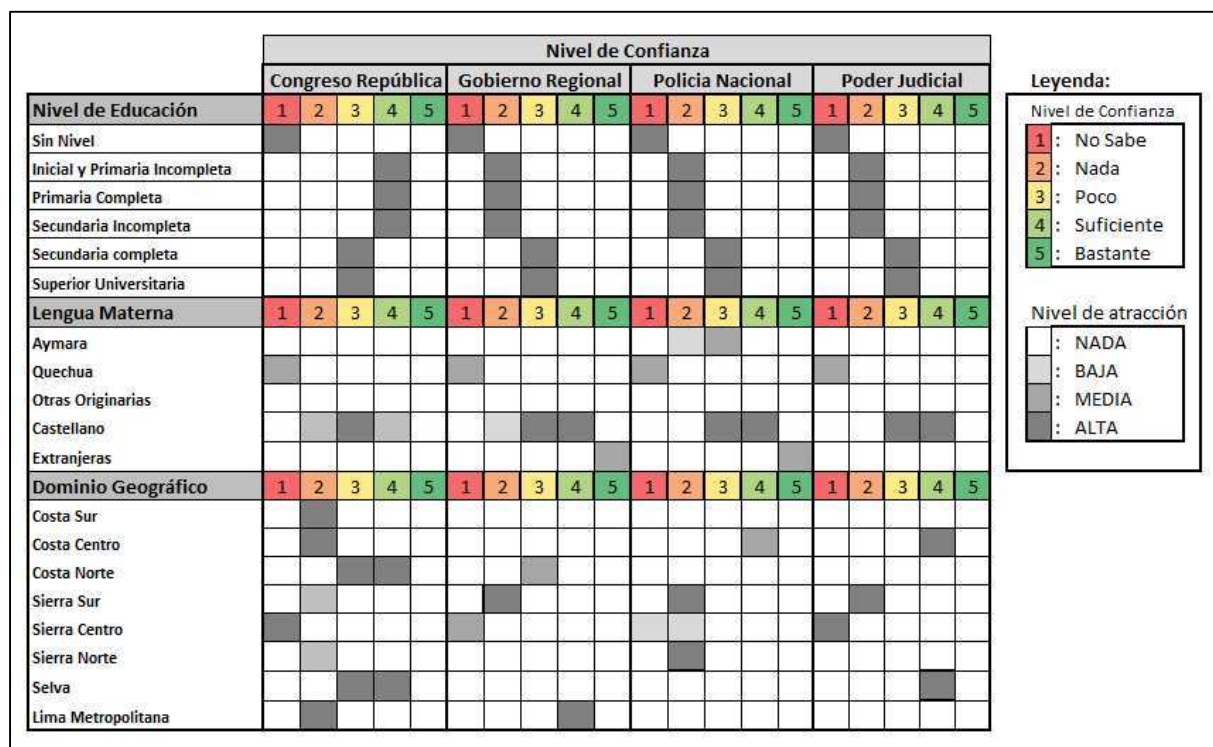


**Ilustración 52:** Dominio Geográfico y la Confianza en el Poder Judicial, 2007, 2013 y 2014



**Ilustración 53:** Material Piso y la Confianza en el Poder Judicial, 2007, 2013 y 2014

### 3.6. Comparativa de comportamiento de variables vs Instituciones



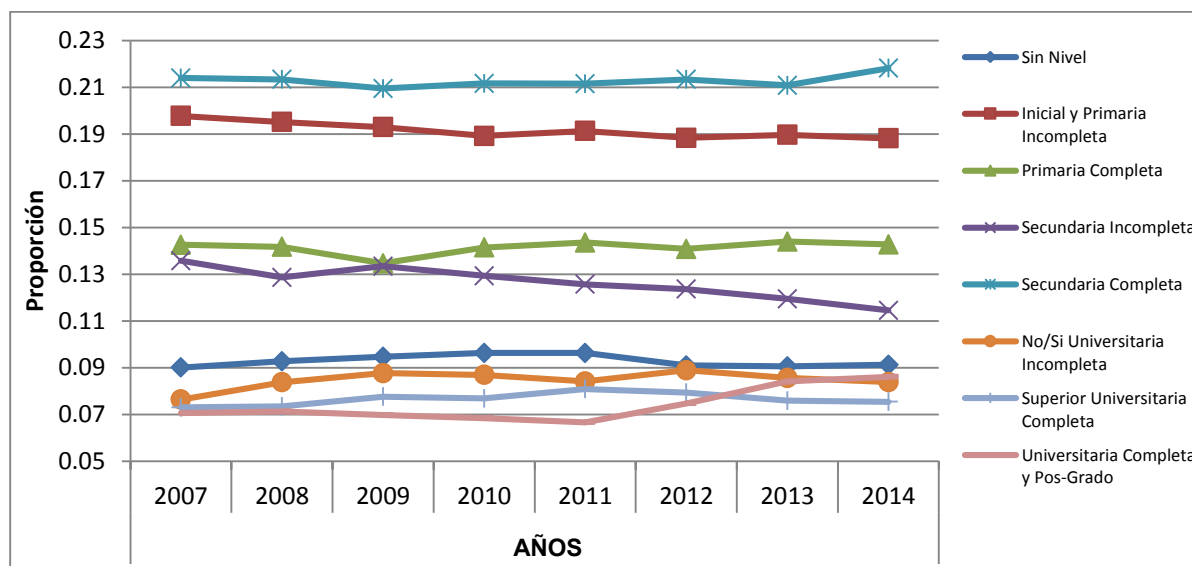
**Ilustración 54:** Comparativa de Comportamiento de las tres principales variables y las Instituciones 2014

En la ilustración 54 se ve una comparativa solo de las 3 variables más influyentes que a la vez son las mismas para las instituciones estudiadas, es construido a base de las atracciones que se visualiza en el análisis de correspondencias dándole un valor significativo en la atracción de modo descriptivo.

En general en el patrón de comportamiento del Congreso de la República es diferente a las demás, tanto que los de la “costa” y “capital Lima metropolitana” tienden a desconfiar como también los más preparados según él [nivel de estudios] y esto no sucede con las demás instituciones, los de [Lengua materna] se ve que los de Lengua “Quechua” están en el sector Rojo(No Sabe) y necesitan urgente atención por parte del gobierno, esto coincide en todas las instituciones estudiadas, también con relación a la variable [Dominio Geográfico] “sierra centro” es el punto de urgencia más relacionado a “No Sabe”.

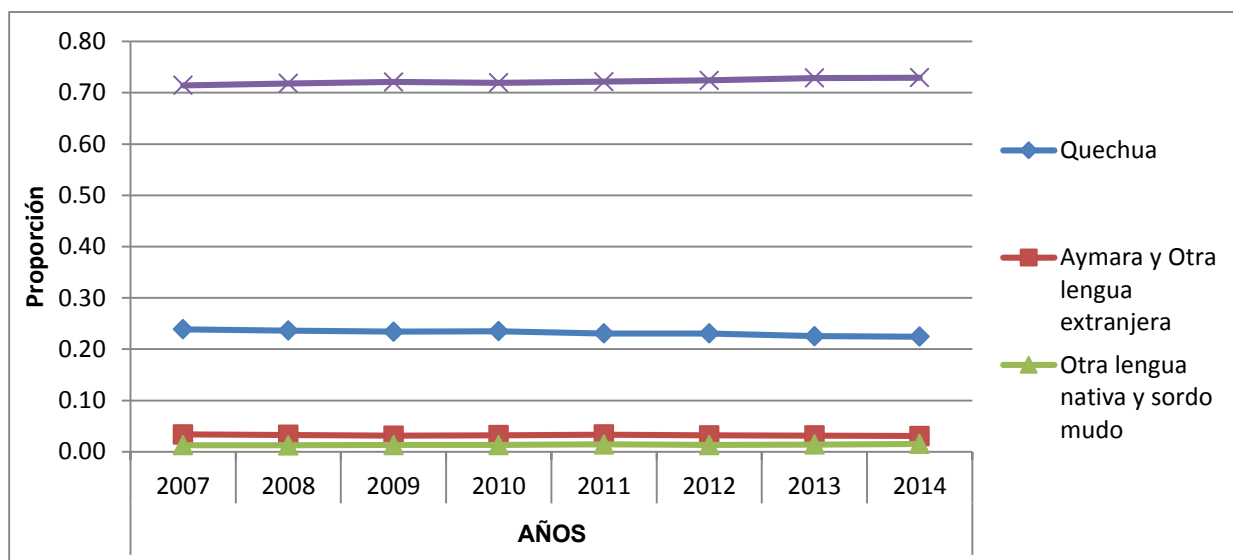
### 3.7. Exploración de las 5 variables más Influyentes

Las 5 variables más influyentes en el 2014, en la confianza con el congreso de la república, tienen un comportamiento de proporciones y cantidades durante el 2007 al 2014, cada variable tiene diferentes valores, los cuales pueden estar relacionados con los niveles de confianza, este último se ve usando técnicas de Análisis de correspondencias.



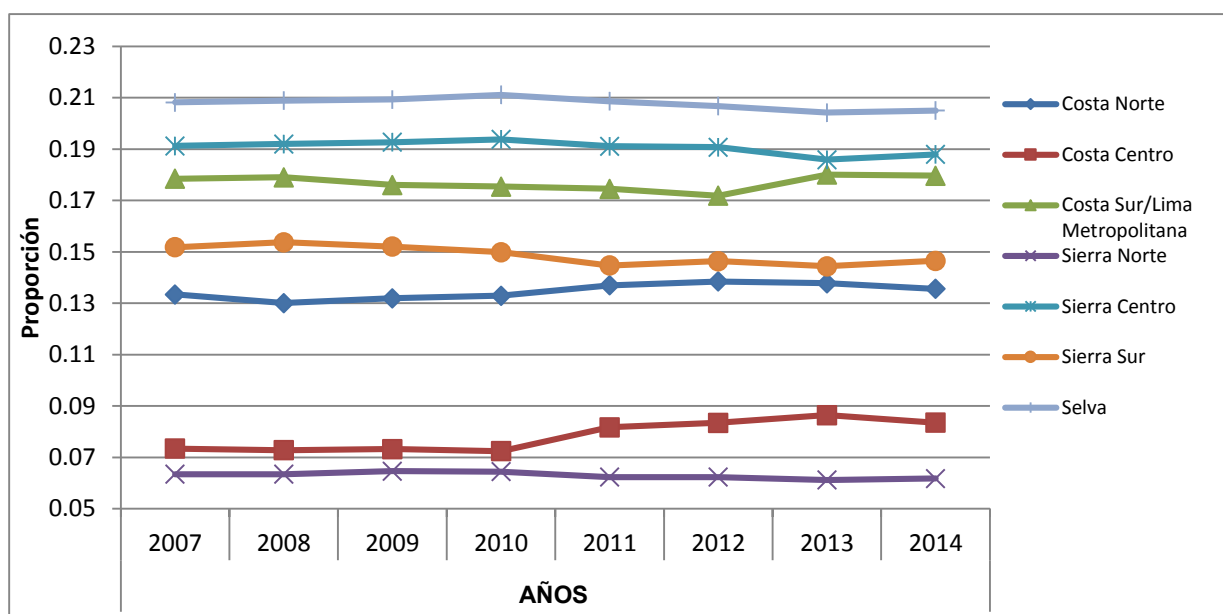
**Ilustración 55:** Evolución Nivel de estudios

En la ilustración 55, el *[Nivel de estudios]* mantiene su proporciones en ciertos niveles, Solo “Secundaria Incompleta” desde el 2009 ha ido bajando, “Nivel Universitario” ha ido aumentando desde el 2011, se ve que “Secundaria completa” tiene mayor proporción en la evolución, le sigue “Inicial” y “Primaria Incompleta”, y los valores bajos son “Superior No Universitario” y “Universitario”.



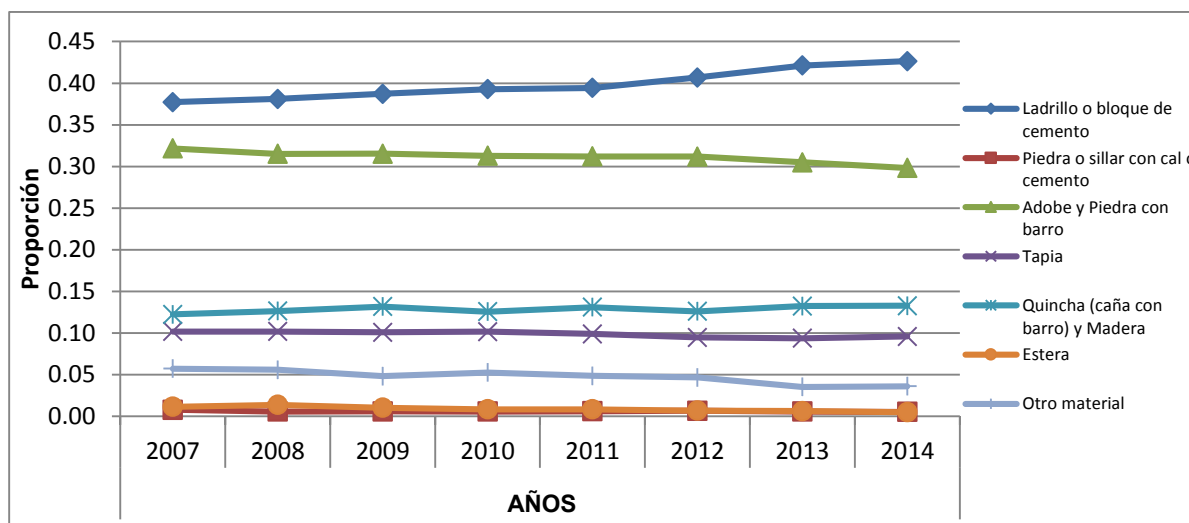
**Ilustración 56:** Lengua Materna

En la ilustración 56, la evolución de *[Lengua Materna]* es estable, esto es que marca la cultura de la población de un determinado origen y es una cualidad que uno aprende en sus primeros años de vida y esto no cambia en el tiempo.



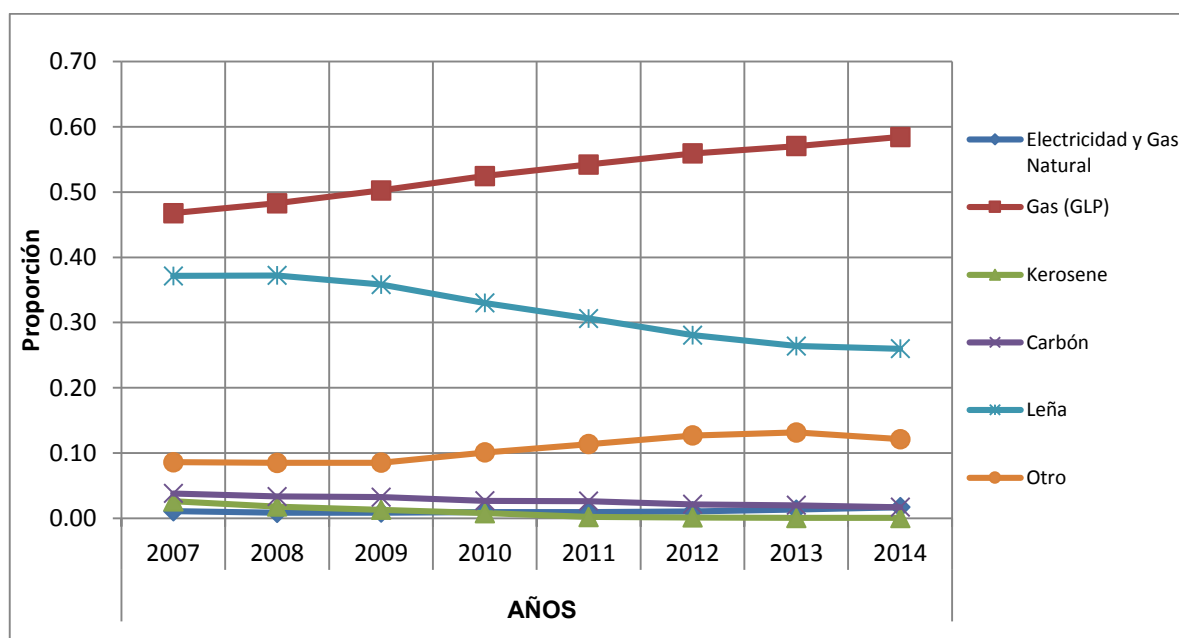
**Ilustración 57:** Evolución Dominio Geográfico

En la ilustración 57, la evolución del *[dominio geográfico]*, se ve los cambios esto que la población de la “Sierra” migran a la “Costa” por un mejor futuro, se ven los incrementos en la “Costa centro y Norte” desde el 2011 y “Costa Sur / Lima Metropolitana” desde el 2012, y en la Sierra baja de proporción.



**Ilustración 58:** Material de Paredes

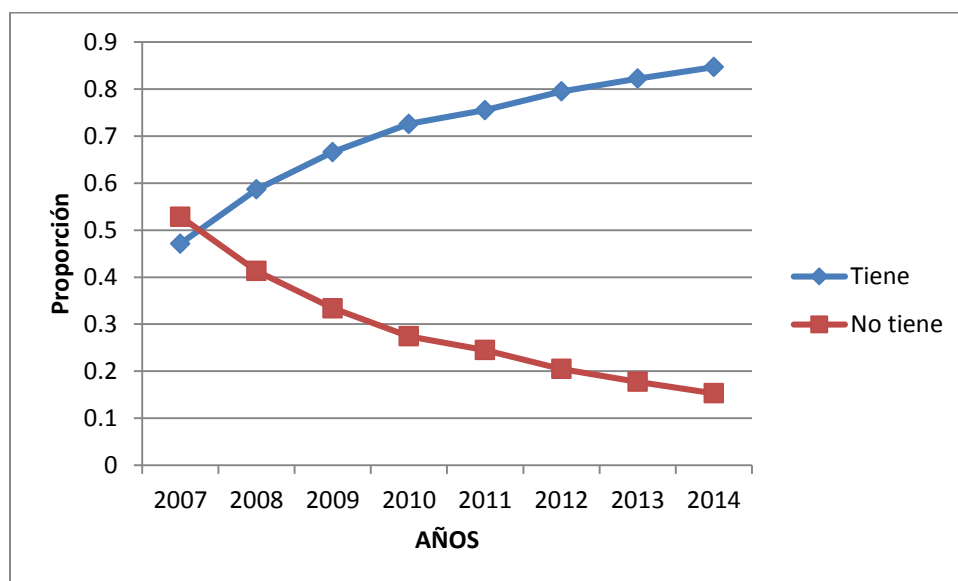
En la ilustración 58, se ve como el [material de paredes] de Bloque y cemento se ha ido incrementando, se sospecha que esto se produce porque cuando se construyen nuevas viviendas estos materiales son los más adecuados, mientras lo demás materiales son viviendas que ya permanecían y las “viviendas de adobe” y “Piedra con barro” se van destruyendo para modernizarse, las casas son el reflejo de la zona cultural o situación económica de las personas.



**Ilustración 59:** Combustible usado en el Hogar

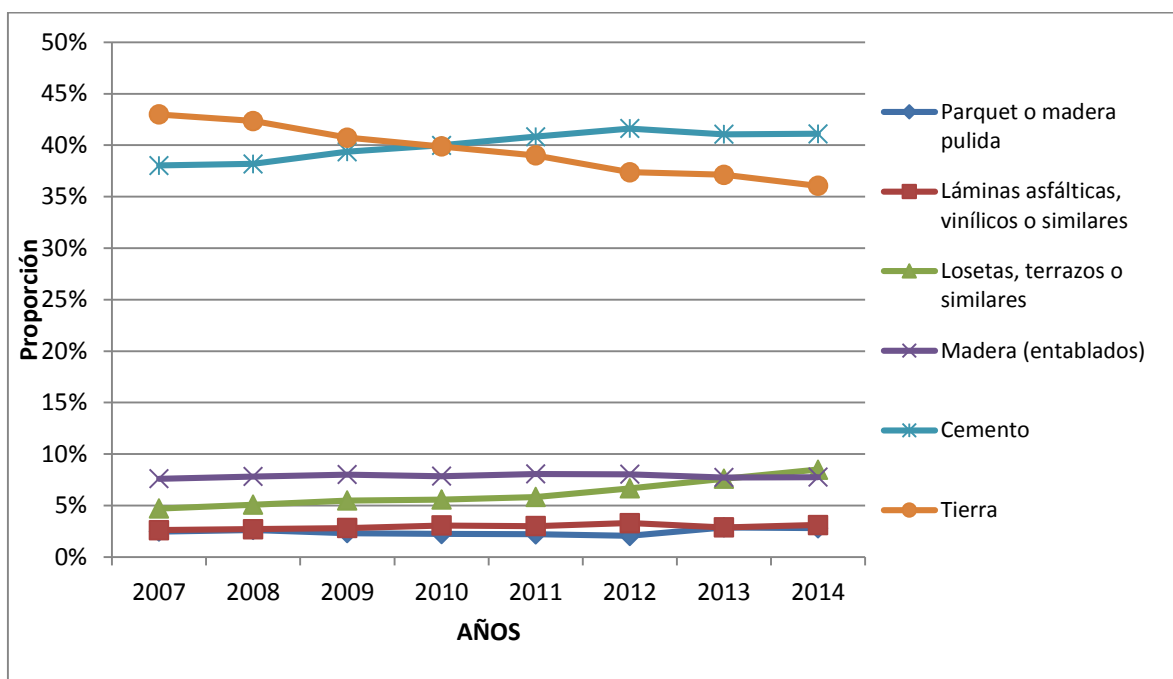
En la ilustración 59 se ve como la electricidad y Gas Licuado de Petróleo (GLP) va aumentando su uso esos últimos años, mientras que la leña va disminuyendo, y otros que se incrementa, si bien en la actualidad casi todas las cocinas y equipos son de Gas (GLP).

Este viene en garrafas y es muy caro, y el Gas Natural aún está en su planificación es por eso los peruanos no tenemos acceso a Gas natural.



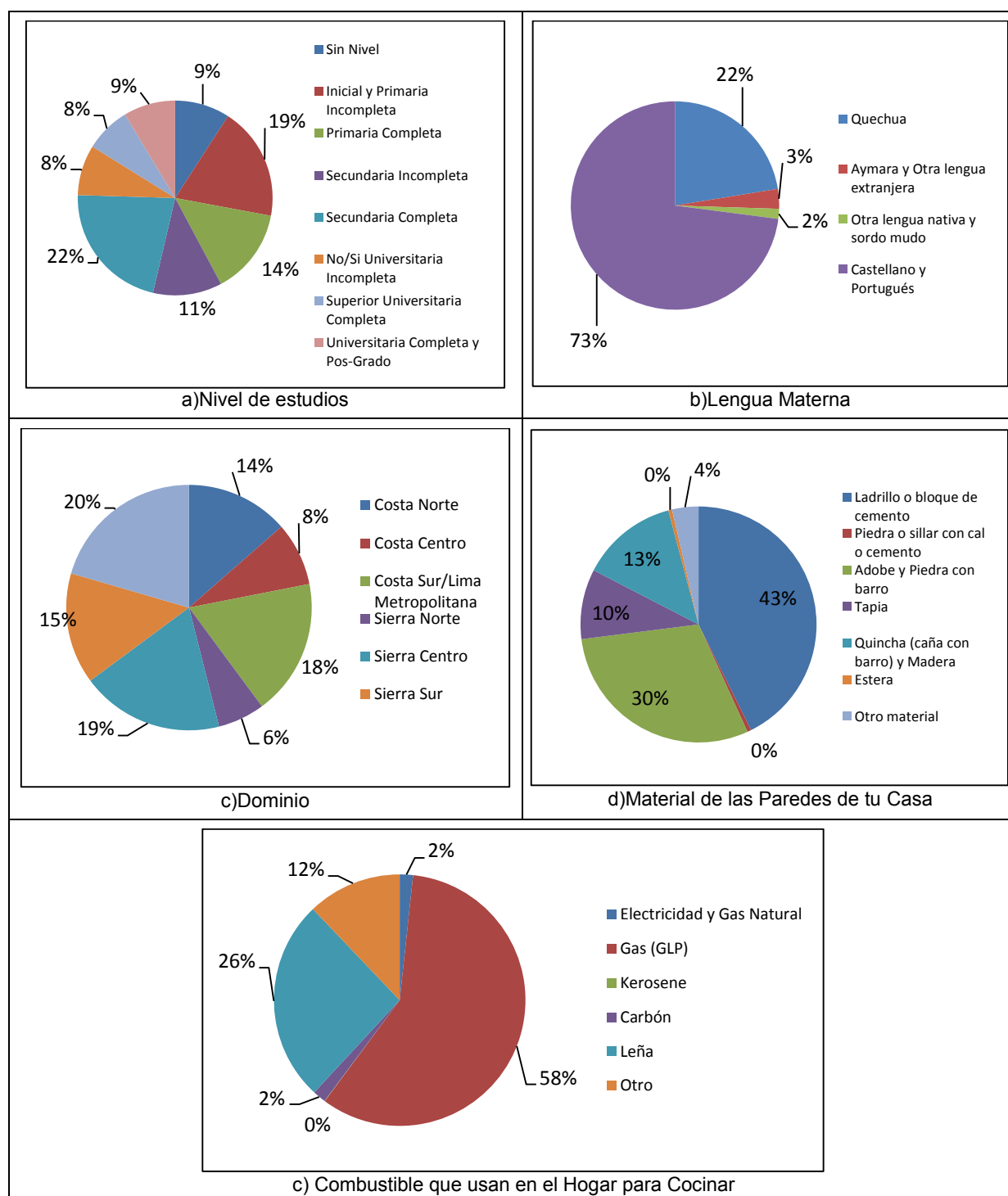
**Ilustración 60:** Acceso a servicios de internet, TV, Cable.

En la ilustración 60, la evolución de *[acceso a servicios de internet, TV, Cable]* esta accesible a más peruanos al pasar los años, esto hace que la población esté más comunicada.



**Ilustración 61:** Material de pisos de la Vivienda

En la ilustración 61, la evolución de *[Material de pisos]* en las viviendas cambia al pasar los años, de tener piso de tierra construyen piso de cemento, y a la vez se van construyendo nuevas viviendas modernas.



**Ilustración 62:** Variables top 5 más influyentes de la base del 2014 con el congreso de la República

En la ilustración 62, se ven las proporciones de las principales variables influyentes del 2014, donde el *[Nivel de estudios]* tiene un alto porcentaje (28%) de personas que a lo sumo tienen nivel primario incompleto. Esto representa a una población sin estudios mínimos que es la cuarta parte de la muestra, mientras que agrupando los estudios superiores desde Superior no universitaria hasta tener posgrado se obtiene un 25%, También el 22% son de secundaria completa, esta es la mayor proporción considerando un segmento solo, con eso sabemos que hay mayor proporción de personas con nivel secundario como también hay gran porcentaje sin Nivel de estudios.

En la *[Lengua materna]* hay desproporción, el más hablado es “Castellano” 73% y le sigue “Quechua” 22% y “Aymara” 3%, en el *[dominio geográfico]*, la “selva” tiene el 20%, “sierra centro” 19 %, “costa sur” y “lima metropolitana” 18%, El *[combustible más usado en el hogar]* con más del 50% es “Gas Licuado de Petróleo” (GLP) con un 58%, le sigue el uso de “Leña” con un 26% y también en un bajo porcentaje esta con un 2% “Gas Natural” y “Electricidad”.



## 4. Conclusiones y Discusiones

### 4.1. Conclusiones

Con relación al objetivo particular uno, se usó satisfactoriamente el proceso de descubrimiento de Conocimiento (KDD), se empezó con la descarga de las bases relacionadas a ENAHO desde el 2007 al 2014 estas en formato de SPSS(.sav), de los repositorios del INEI-PERÚ, luego se diseñó una base de datos en SQL Server donde se cargó las bases de las encuestas, se continuo la limpieza, transformación, integración y selección de datos hasta finalizar con la obtención de una tabla Analítica, esta permitió la exploración de las variables para la selección de variables potenciales influyentes.

En la base se encontraron datos de 13 instituciones que permanecieron estables del 2007 al 2014, referidos a la confianza, se realizó el análisis con datos del 2014, se seleccionó el top 4 de instituciones con mayor proporción en “Nada de confianza”, como resultado se obtuvo que el Congreso de la República es la institución con mayor desconfianza, le siguen Gobierno regional, Poder Judicial y Policía Nacional en ese orden, las proporciones de desconfianza permanecen estables en el tiempo del 2007 al 2014.

Con relación al objetivo general, se obtuvieron 5 variables más influyentes, mediante el algoritmo de árbol de decisión, en orden de influencia son: *[Nivel de estudios]*, *[Lengua Materna]*, *[Dominio Geográfico]* siendo estas 3 los mismos para las 4 instituciones, y los 2 restantes son iguales para todas las instituciones, resultando relevantes *[Material de paredes de la vivienda]* y *[Material de Pisos de la vivienda]*, *[El tipo de combustible que usan para cocinar]*, *[Cuenta con Servicios de Internet, TV, Cable]*. Se determinan los valores de influencias mediante el indicador de chi cuadrado de contingencia desde el 2007 al 2014 muestran valores que no superan 0,5 siendo una relación débil.

En relación al segundo objetivo particular, se realizó el análisis de la evolución de la influencia desde el 2007 al 2014, donde se calculó una matriz de coeficiente de contingencia chi-cuadrado de las variables, los resultados muestran que el *[nivel de estudios]* se mantiene estable en su posición de influencia para las 4 instituciones, *[Lengua materna]* tiene variaciones de posiciones y desde el 2012 al 2014, pero mantiene su posición en las 4 instituciones y solo mantiene estable en relación con la Policía nacional, las demás variables muestran muchas variaciones en el periodo analizado.. La variable *[el hogar no tiene Teléfono Fijo, Cable, Internet, etc]*, ha ido descendiente su influencia en los años del 2008 al 2014.

Las atracciones de influencia son un análisis más detallado y complementario al de asociación e influencia, donde las modalidades de dos variables son atraídas, este análisis se hizo mediante el análisis Factorial de Correspondencias. De este análisis se desprende entre otras conclusiones que hay más desconfianza en el Congreso de la república cuando se tiene mayor *[nivel estudios]*

Con relación al tercer objetivo particular, se realizó el análisis de temporalidad de las variables más influyentes, como resultado algunas variables mantienen su proporcionalidad y otras no, el *[nivel educación]* solo el “Nivel Universitario” ha ido aumentando desde el 2011 permaneciendo casi constante los demás niveles, *[lengua Materna]* permanece invariable, el *[dominio geográfico]* se ve que los de la “Sierra” migran a la “Costa”, esto por los incrementos en la “Costa centro y Norte” desde el 2011 y “Costa Sur / Lima Metropolitana” desde el 2012, y en la “Sierra” va disminuyendo la población, el *[material de paredes]* de “Bloque” y “cemento” se ha ido incrementando, mientras material de “adobe” y “Piedra con barro” van, *[Combustible usado en el Hogar]* la “electricidad” y “Gas Licuado de Petróleo” (GLP) van incrementando su consumo durante los años, mientras que la leña va disminuyendo, el *[acceso a servicios de internet, TV, Cable]* la accesibilidad va aumentando en los años. *[Material de pisos]* en las viviendas cambian al pasar los años, de tener piso de tierra a piso de cemento.

## 4.2. Discusiones

La confianza en las instituciones del gobierno es muy importante, tal que este trabajo busca determinar un factor muy influyente, visto que el Nivel de estudios es un factor que influye, los resultados fueron contrastados mediante Arboles de decisión, Chi cuadrado, análisis de Correspondencias.

El gobierno debe de impulsar un programa social de fortalecimiento en las personas que no han tenido la oportunidad de tener algún estudio, que ellos de alguna forma tienen un desconocimiento sobre las instituciones del gobierno o por su nivel de estudios no son comprendidos por los funcionarios de las instituciones, esto hace que estos sientan una desconfianza.

Los métodos de minería de datos muy usados en campos de finanzas y marketing, estos métodos son adaptados a la parte social de Gobierno visto en este trabajo, no se asegura que estos resultados puedan servir a otros países

## 5. Trabajos Futuros

Este trabajo sirve como principal documento para realizar futuros planes de campañas para fortalecer la confianza con las instituciones, si bien se realizaron con las 4 principales desconfiadas se podría repetir el procedimiento para los 9 restantes.

Si bien se eligió los 5 principales factores para el estudio, se puede repetir el procedimiento y estudiar las demás variables, el resultado te da la perspectiva de cómo se comportan las variables influyentes en el tiempo, puede ser aprovechado para iniciar planes de Campañas de fortalecimientos en la confianza.

El Perú tiene 3 idiomas más hablados estos son el Castellano, Quechua y Aymara, estos dos últimos son los más afectados y atraídos a No sabe de las Instituciones, estos resultados se pueden usar para la planificación de alguna Campaña de Concientización y que funcionarios públicos lleguen en el mismo idioma en los diferentes canales de comunicación, Televisión, Radio, etc. al segmento afectado. Luego de esto se podría repetir el estudio para analizar si el impacto de la campaña fue positiva.

El instituto Nacional de estadísticas desarrolla proyectos de Censos y encuestas todos los años, estos datos están disponibles para cualquier investigador, esto es materia prima para buscar conocimientos que podrían servir para mejorar los problemas de la sociedad.

## 6. Referencias Bibliográficas

- [1] P. V. Britos, A. Hossian, R. García Martínez y E. Sierra, *Minería de Datos basada en Sistemas Inteligentes*, Buenos Aires: Nueva Librería S.R.L., 2005, pp. 38-39.
- [2] «Instituto Nacional de Estadística e Informática,» 10 Abril 2015. Disponible: <http://www.inei.gob.pe/bases-de-datos/>.
- [3] Pnud, «La protesta social es parte de la democracia,» 28 Noviembre 2012.
- [4] Latinobarómetro, «Informe 2013,» Santiago de Chile, 2013.
- [5] Diario La República, «Latinobarómetro,» 3 Noviembre 2013. Disponible: <http://archivo.larepublica.pe/columnistas/virtu-e-fortuna/latinobarometro-2013-03-11-2013>. [Último acceso: 12 Diciembre 2014].
- [6] N. Liendo, «El crecimiento económico y la confianza en las instituciones de gobierno: Un análisis comparado de la evolución en las percepciones de las élites y los ciudadanos de América Latina,» *Boletín Pnud&Instituto de Iberoamerica*, 2011.
- [7] C. Moreno y E. Lépoire, «CONFIANZA PÚBLICA, PARTICIPACIÓN CIUDADANA Y APEGO A LA LEY EN LA ARGENTINA POSTDEVALUACIÓN,» Universidad Católica Argentina. , Buenos Aires, 2007.
- [8] J. Aguilar, *LA PERCEPCIÓN DE LA SEGURIDAD Y LA CONFIANZA EN LAS INSTITUCIONES PÚBLICAS*, Instituto Universitario de Opinion Publica (IUDOP), 2013.
- [9] Camara de Diputados, «Encuesta telefónica sobre confianza en las instituciones,». Disponible: <http://www5.diputados.gob.mx/index.php/camara/Centros-de-Estudio/CESOP/Opinion-Publica/Encuestas/Encuesta-telefonica-sobre-confianza-en-las-instituciones>. [Último acceso: 05- 03 -2014].
- [10] J. Han y M. Kamber, *Data Mining Concepts and Techniques*, Second Edition ed., Boston: El Sevier, 2006.
- [11] E. Hernández Arrollo, *Manual de estadística / Handbook of Statistics*, Bogotá: Universidad cooperativa de Colombia, 2006.
- [12] IBM, *IBM SPSS Statistics Base 22*. Disponible: [http://www.sussex.ac.uk/its/pdfs/SPSS\\_Statistics\\_Base\\_22.pdf](http://www.sussex.ac.uk/its/pdfs/SPSS_Statistics_Base_22.pdf).

- [13] M. Greenacre, *La práctica de análisis de correspondencias*, Primera edición ed., Bilbao: Rubes, 2008.
- [14] E. Alpaydin, *Introduction to machine Learning*, Segunda ed., London: Adaptive Computation, 2010, pp. 185-208.
- [15] IBM, *IBM SPSS Decision Trees 21*, 2012. Disponible:  
[http://www.sussex.ac.uk/its/pdfs/SPSS\\_Decision\\_Trees\\_21.pdf](http://www.sussex.ac.uk/its/pdfs/SPSS_Decision_Trees_21.pdf)
- [16] F. Sagasti, P. Patron, M. Hernández y N. Lynch, *DEMOCRACIA Y BUEN GOBIERNO, HACIA LA GOBERNABILIDAD DEMOCRÁTICA EN EL PERÚ*, Tercera ed., Lima, Lima: Grafica Pacific Press S.A., 1999.
- [17] H. Naessens, *Ética Pública y transparencia*, XIV Encuentro de Latinoamericanistas Españoles ed., México, 2010.
- [18] C. Moreno y E. Lépre, «CONFIANZA PÚBLICA, PARTICIPACIÓN CIUDADANA Y APEGO A LA LEY EN LA ARGENTINA POSTDEVALUACIÓN,» Universidad Católica Argentina. , Buenos Aires, 2007.
- [19] D. Montalvo, «Perspectivas desde el Barómetro de las Americas 2010,» *Vanderbilt University*.
- [20] N. Liendo, «El crecimiento económico y la confianza en las instituciones de gobierno: Un análisis comparado de la evolución en las percepciones de las élites y los ciudadanos de América Latina,» *Boletín Pnud&Instituto de Iberoamerica*, 2011.
- [21] C. Gómez Díaz de León y X. Arango Morales, *Administración Pública y Gobernanza: El Papel de la Gestión de la Ética en las Organizaciones Públicas*, Daena: Internacional Journal of Good Conscience, 2012.
- [22] C. Enrique Trejo, *ÉTICA Y POLÍTICA: CONSTRUCCIÓN DE LA CONFIANZA EN LAS INSTITUCIONES PÚBLICAS*, 2006.
- [23] N. Douglass C., *INSTITUCIONES CAMBIO INSTITUCIONAL Y DESEMPEÑO ECONÓMICO, FONDO DE CULTURA ECONÓMICA DE MÉXICO*, México, 1990.
- [24] A. de Jesús Suárez, *Deudas y progresos sociales en un país que hace frente a su bicentenario. Argentina 2004 - 2010*, b. d. l. d. s. argentina, Ed., 2011.
- [25] J. Aguilar, *LA PERCEPCIÓN DE LA SEGURIDAD Y LA CONFIANZA EN LAS INSTITUCIONES PÚBLICAS*, Instituto Universitario de Opinión Pública (IUDOP), 2013.
- [26] Gobierno de los Estados Unidos Mexicano, «Plan Nacional de Desarrollo,» 2007. Disponible:  
[http://pnd.calderon.presidencia.gob.mx/pdf/PND\\_2007-2012.pdf](http://pnd.calderon.presidencia.gob.mx/pdf/PND_2007-2012.pdf). [Último acceso: 12 Abril 2015].

- [27] Cepal, «INSTITUCIONES,» Disponible:  
<http://www.cepal.org/prolider/noticias/paginas/5/31975/INSTI.pdf>. [Último acceso: 3 Abril 2015].
- [28] T. Mitchell, Machine Learning, C. S. Series, Ed., McGraw-Hill International Editions, 1997, p. 52–78.
- [29] t. Pang-Ning, S. Michael y K. Vipin, Introduction to Data Mining, Boston: Pearson Adisson Wesley.
- [30] Johnson y Wichern, Applied Multivariate Statistical Analysis, sexta edición ed., 2007, pp. 671-701.

## 7. ANEXO

Tabla 12: Módulos sobre la encuesta nacional de hogares ENAHO 2014

Nro	Año	Período	Código Encuesta	Código Módulo	Módulo
1	2014	55	440	1	<a href="#">Características de la Vivienda y del Hogar</a>
2	2014	55	440	2	<a href="#">Características de los Miembros del Hogar</a>
3	2014	55	440	3	<a href="#">Educación</a>
4	2014	55	440	4	<a href="#">Salud</a>
5	2014	55	440	5	<a href="#">Empleo e Ingresos</a>
6	2014	55	440	7	<a href="#">Gastos en Alimentos y Bebidas (Módulo 601)</a>
7	2014	55	440	8	<a href="#">Instituciones Benéficas</a>
8	2014	55	440	9	<a href="#">Mantenimiento de la Vivienda</a>
9	2014	55	440	10	<a href="#">Transportes y Comunicaciones</a>
10	2014	55	440	11	<a href="#">Servicios a la Vivienda</a>
11	2014	55	440	12	<a href="#">Esparcimiento , Diversión y Servicios de Cultura</a>
12	2014	55	440	13	<a href="#">Vestido y Calzado</a>
13	2014	55	440	15	<a href="#">Gastos de Transferencias</a>
14	2014	55	440	16	<a href="#">Muebles y Enseres</a>
15	2014	55	440	17	<a href="#">Otros Bienes y Servicios</a>
16	2014	55	440	18	<a href="#">Equipamiento del Hogar</a>
17	2014	55	440	22	<a href="#">Producción Agrícola</a>
18	2014	55	440	23	<a href="#">Subproductos Agrícolas</a>
19	2014	55	440	24	<a href="#">Producción Forestal</a>
20	2014	55	440	25	<a href="#">Gastos en Actividades Agrícolas y/o Forestales</a>
21	2014	55	440	26	<a href="#">Producción Pecuaria</a>
22	2014	55	440	27	<a href="#">Subproductos Pecuarios</a>
23	2014	55	440	28	<a href="#">Gastos en Actividades Pecuarias</a>
24	2014	55	440	34	<a href="#">Sumarias ( Variables Calculadas )</a>
25	2014	55	440	37	<a href="#">Programas Sociales (Miembros del Hogar)</a>
26	2014	55	440	77	<a href="#">Ingresos del Trabajador Independiente</a>
27	2014	55	440	78	<a href="#">Bienes y Servicios de Cuidados Personales</a>
28	2014	55	440	84	<a href="#">Participación Ciudadana</a>
29	2014	55	440	85	<a href="#">Gobernabilidad, Democracia y Transparencia</a>

Fuente. INEI (Instituto Nacional de estadísticas e informática)